Speech Fluency and Ageing

A study of disfluencies, speech errors and speech rate in three age groups

Harry Polfeldt

Abstract

Several physiological changes in the speech production process in old age have been documented, yet studies of cognitive changes are few and inconclusive. The current study investigates changes in rates of disfluencies, rates of speech errors and speech rate to examine if speech fluency changes throughout the adult lifespan, as well as whether the influence of task complexity on fluency remains constant regardless of the age of the speaker. Elicited speech from six young, six middle-aged and five older speakers of Swedish describing five simple and five complex stimuli is analysed for disfluencies, speech errors and speech rate. The study finds no significant relationship between the speaker's age and rate of disfluencies, speech errors or speech rate, and no cohort is significantly more affected by task complexity. The findings suggest that fluency remains relatively intact throughout the adult lifespan.

Acknowledgements

Table of Contents

1. Introduction

Older adults are becoming an increasingly large part of the global population, growing faster than any other age group (Hooper & Craildis, 2009). In Sweden, the number of individuals over the age of 80 is expected to increase by 50% between 2018 and 2028 (SCB, 2018). Although aging is associated with many physiological and cognitive changes, such as muscle atrophy (Volpi et al., 2004) and worsened memory (Sandson et al., 1987), it is not clear whether language production abilities remain intact in old age. While usually considered to be declining with age (Gollan & Goldrick, 2019), most documented changes in the language production system are physiological, such as a reduced maximum phonation time (Mueller, 1982, referred to in Searl et al., 2002), whereas the cognitive aspects of language production, such as speech fluency, have been studied markedly less in relation to aging (Searl et al., 2002). If speech fluency is affected by age, it might entail that certain aspects of the cognitive speech production process change throughout the lifespan. Since older adults constitute a progressively larger segment of the population, it is vital for speech-language pathologists to be able to differentiate and make diagnostic distinctions between normal and pathological speech processes in relation to aging (Searl et al., 2002). Moreover, while it is known that task complexity negatively affects fluency (Bortfeld et al., 2001), no study seems to have examined whether task complexity affects speakers of different ages in different ways. If task complexity disproportionately affects speakers of a certain age, speech pathologists must be aware of the potential influence of this factor to create balanced tests.

The current study aims to examine whether fluency changes with age. To probe this question, the study investigates three main dependent variables. First of all, disfluencies, defined as interferences in the otherwise fluent flow of speech, arise from errors or delays in the cognitive speech production process and are consequently relevant to general theories of language production (Fraundorf et al., 2014) as the type and location of disfluencies can reveal insights into the problems we encounter when producing speech (Lickley, 2015). Second, speech errors, which are mismatches between what a speaker intends to say and what they actually produce (Harley, 2006), reveal aspects of lexical selection (i.e. choosing what word to say), and associations between mental concepts during the production of language (Levelt, 1989, p. 234). Third, speech rate, or the speed at which a speaker produces their words, is correlated with the rate of disfluencies and speech errors (Levelt, 1989, p. 355-359;

368), and is also partially a reflection of difficulties in lexical retrieval (i.e. finding the target word in the mental lexicon) (Horton et al., 2010).

2. Background

Disfluencies, pauses, speech errors and speech rate are connected to the production of language. Consequently, to study them and their implications, an understanding of the speech production process is necessary. By detailing the process involved in conceptualising, formalising, articulating and monitoring speech, it becomes clear how errors in the speech production appear as well as what they entail. This section presents a simplified summary of Levelt's (1989) model of how speech is produced and then treats disfluencies, pauses, speech errors and speech rate and task complexity in the context of speech production. Finally, previous studies on fluency and aging are summarised.

2.1. Speech Production

2.1.1. Levelt's Speech Production Model

One of the most influential models concerning the speech production process is that of Levelt (1989). Levelt's (1989) model of speech production claims that speech consists of several processing components which work in parallel. To produce speech, the conceptualizer produces preverbal messages, which serve as the input to the formulator which in turn produces articulatory plans, which serve as the input to the articulator which in turn outputs overt, spoken speech. Overt speech is then monitored through audition (i.e. listening) and then processed through the speech-comprehension system, the same system involved in processing and comprehending the speech of other interlocutors.

The first processing component is the conceptualizer. This stage is entirely preverbal (Levelt, 1989, p. 9), effectively formulating the basic concept of what the speaker wishes to express. To conceptualise a message, the speaker relies on knowledge stored in their memory as well as current situational awareness such as who they are speaking with (Levelt, 1989, p. 10). The output of conceptualisation is the preverbal message which acts as the input to the formulator, the second processing component (Levelt, 1989, p. 9). The formulator transforms the preverbal message into a phonetic plan through grammatical encoding and phonological encoding. During grammatical encoding, the formulator accesses lemmas (Levelt, 1989, p. 11), namely non-phonological information about a lexical item (Levelt, 1989, p. 6) such as

the semantic meaning and the syntactic properties (i.e. being a noun, a verb, etc). Lemmas are then ordered into syntactic phrases and subphrases (Levelt, 1989, p. 11). During phonological encoding, the formulator accesses the mental lexicon and retrieves the morphology, such as different affixes or conjugations, and phonology, namely the sequence of sounds, stress, etc, for each lemma. The result is a phonetic plan for how the speech will eventually be produced by the physical speech production system (Levelt, 1989, p. 12). The phonetic plan then acts as input to the third component, the articulator, which uses physical musculature to produce overt, spoken speech (Levelt, 1989, p. 11). The fourth and final part of the speech production process is the speech-comprehension system, which consists of various subcomponents (Levelt, 1989, p. 13). These components monitor both the speaker's planned internal speech and their overtly produced speech, controlling for errors and other issues in the message (Levelt, 1989, p. 14).

2.1.2. Disfluencies and Unfilled Pauses

The speech production process presented above is one of a human's most complex skills from a cognitive and motoric perspective (Levelt, 1989, p. XIII). Due to its complexity, problems in the speech production process are common. Disfluencies appear as frequently as up to 10 times per 100 words in healthy speech (Adams, 1980; Guitar, 1998, referred to in Searl et al., 2002). These disfluencies take many different forms, of which some common types are discussed in this section.

First of all, it is important to understand why a speaker pauses to plan in the first place. One reason is that during conceptualisation, the speaker might simply not know what to say or how to say it. This is in turn influenced by the length and structure of the utterance being planned (Lickley, 2015), with more complex messages needing more time to plan and consequently eliciting more disfluencies (Goldman-Eisler 1968; Good & Butterworth 1980; Maclay & Osgood 1959, referred to in Lickley, 2015). Moreover, the speaker might pause to search their memory for information relevant to conceptualising the message (Lickley, 2001, referred to in Lickley, 2015). A speaker might also need to pause and plan during lexical retrieval (Lickley, 2015), namely at the formulation stage, where longer and less frequent words are harder to retrieve and consequently associated with more planning and hesitations (Hartsuiker & Notebaert, 2010; Severens et al., 2005, referred to in Lickley, 2015). It is also possible that the articulatory complexity of a word affects fluency, with more complex words

requiring more planning (Mooshammer et al., 2012, referred to in Lickley, 2015). In sum, hesitations act as overt and measurable indications of activity in processing (Chafe, 1985, referred to in Kendall, 2013, p. 24).

As for the specific types of hesitations and disfluencies, one notable type is unfilled pauses. These are intervals of silence in speech (Matthews, 2007). However, not all silences are disfluencies. A common type is delimatitive pauses, which correlate with punctuation (Warren, 2013, p. 18) or intonation (Levelt, 1989, p. 308) to assist the listener in structuring the message into clauses and phrases (Warren, 2013, p. 223). Delimitative pauses play intricate roles in conversations, for instance indicating the end of a speaking turn (Levelt, 1989, chap. 3). Another type of pause is physiological pauses, namely pausing for breathing. This type of pause is largely disregarded in psycholinguistic research. A different type of pause is an articulatory pause, that is, inevitable pauses shorter than 200 milliseconds which occur during articulation, for instance during the occlusive phase of a plosive (Warren, 2013, p. 18), such as when the lips close to build up air pressure during the production of a /p/ or /b/. For a silent pause to be counted as a disfluency, it must be subjectively considered to interrupt an ongoing unit of speech (Levelt, 1989, p. 33). For instance, while unfilled pauses are more common at the end of a clause or a phrase (Levelt, 1989, p. 391), often matching sentence boundaries (Levelt, 1989, p. 386), pauses are less common between function words, such as *a* or *the,* and content words, such as *dog* or *run.* In such instances, the unfilled pause might be considered to interrupt an ongoing unit, subjectively not occurring in a natural position. In fact, unfilled pauses as a disfluency might hamper the listener's comprehension of the message (Warren, 2013, p. 7). Consequently, an unfilled pause can be defined as a period of silence longer than 200 milliseconds that interferes with an ongoing unit. If a pause does not interrupt a unit, it might be delimitative or perhaps simply due to difficulty in planning or retrieving upcoming words (MacGregor et al., 2010).

Unfilled pauses as a disfluency can signal many different issues based on their location in the message (Warren, 2013, p. 2). In general, unfilled pauses occur when no new lemmas are available for phonological encoding, either due to issues in conceptualisation or grammatical encoding, which means the speaker has to pause for planning (Levelt, 1989, p. 386-387). Moreover, unfilled pauses seem to be closely related to syntactic planning (Rose, 2017), since more syntactically complex sentences include more and longer pauses due to the increased need for planning (Warren, 2013, p. 29). However, while syntactic planning might be the

main cause of unfilled pauses, the disfluency type can also occur in relation to issues in lexical retrieval (Levelt, 1989, p. 33; Warren, 2013, p. 40), which is part of the formulation stage, or due to more general conceptualization of the message (Levelt, 1989, p. 386; Rose, 2017). In sum, unfilled pauses can act as indicators of cognitive load (Cappella, 1985, referred to in Kendall, 2013, p. 24).

In contrast, filled pauses are articulated sounds that do not convey any propositional meaning (Mahl, 1956; Clark & Fox Tree, 2002, referred to in Rose, 2017), meaning that filled pauses do not convey falsifiable statements, unlike the majority of language. While filled pauses vary phonologically across languages (Levelt, 1989, p. 483), *eh* and *uhm* are the most common in English (Clark & Fox Tree, 2002).

According to Clark & Fox Tree (2002), there are three main views on the purpose of filled pauses. One view is the *filler-as–word* view, which proposes that filled pauses are interjections on par with true words such as *Aha* or *Damn* (Clark & Fox Tree, 2002, p. 77). Another is the *filler-as-nonlinguistic signal* view, which proposes that filled pauses have communicative functions such as holding the floor (i.e. maintaining control of the speech turn) (Clark & Fox Tree, 2002), or edit signals which inform the listener that an error in the speech has been detected in the speech comprehension process by the speaker and that a repair is underway, meaning the utterance will be remedied or replaced (Levelt, 1989, p. 482). The third view is the *filler-as-symptom* view, which sees filled pauses as an automatic consequence of a problem in the speech production process. In this view, filled pauses are merely hesitations used to provide the speaker time to plan their message (Batliner et al., 1995). They lack communicative intent and meaning, and are simply a symptom of issues in the speech production process (Levelt, 1989, p. 484). Filled pauses are usually not regarded as associated with problems at any specific point in the speech production process (Rose, 2017), yet some evidence suggests that filled pauses are more related to issues in lexical retrieval than anything else (Beattie & Butterworth, 1979, referred to in Rose, 2017; Warren, 2013, p. 77).

Furthermore, interruptions occur when the speaker abruptly terminates their original utterance (Warren, 2013, p. 75) due to noticing an error (Levelt, 1989, p. 459), either in overt or internal speech (Levelt, 1989, p. 478). Interruptions are part of repairs, when what was said or was going to be said is incorrect, or revisions, where the message was in some manner

incomplete or inappropriate (Warren, 2013, p. 76). Since speech is constantly monitored through the speech comprehension process (Warren, 2013, p. 72), speakers will interrupt themselves immediately upon detecting an error, referred to as the main interruption rule (Levelt, 1989, p. 478). However, while most interruptions occur directly after the error in question, around one third do not occur until at least one word after the error (Levelt, 1989, p. 479-480). Interruptions are also often followed by an edit signal (Warren, 2013, p. 76).

Moreover, prolongations, as described by Betz & Wagner (2016), are noticeable lengthenings of one or more phonemes, rendering the duration of the word longer than usual. Consequently, speech rate is decreased and the listener may perceive the prolongation as a noticeable hesitation. The prolongation is utilised by the speaker to gain time for planning, and also informs the listener that the message is still being planned and is thus not yet complete, which holds the floor for the speaker. However, according to Betz & Wagner (2016), prolongations are the least disfluent hesitation type since speech never truly stops, and listeners will perceive a prolongation as less disruptive than a filled pause. Similar to other hesitations, prolongations can signal a variety of problems in the speech production process, and are not associated with any specific component.

Next, repetitions occur when the speaker repeats an already produced segment of speech. While repetitions are closely associated with stuttering (Lickley, 2018), they are also one of the most common disfluencies in normal speech (Lickley, 2015). In stuttering-like repetitions, there is muscular tension, the repeated segment is usually a single phoneme or a small cluster of phonemes, the repetition often occurs more than once, and the repetition does not arise due to problems in planning as the speaker is fully aware of what they wish to express. In contrast, a normal disfluent repetition consists of at least one syllable, but often of multisyllabic words or phrases, is usually not repeated more than once, entails no significant increase in muscular tension, and arises due to the need to pause for planning during conceptualisation and formulation. Regardless of the number of times a segment is repeated, it is usually considered a single instance of a repetition (Bortfeld et al., 2001; Johnson, 1961; Leeper & Culatta, 1995). Moreover, there are also fluent repetitions, such as the rhetorical repetition of a word for emphasis. Fluent and disfluent repetitions differ phonologically since fluent repetitions fit the prosodic pattern of the phrase whereas disfluent repetitions break the expected prosody and are often accompanied by unexpected pauses (Lickley, 2018).

Lastly, it should be noted that previous studies of disfluencies and aging vary significantly in what they define as a disfluency. Older studies, such as Duchin & Mysak (1987), Leeper & Culatta (1995) and Searl et al. (2002), typically rely on modified versions of Johnson's (1961) disfluency types. In his influential study, Johnson (1961) studies speech thought to be representative of, or related to, disfluency in spontaneous, semi-spontaneous and controlled speech of 200 college-aged adult males and females. Creating his own categorization system based on the findings, he calculates the prevalence of each type of disfluency per 100 words. The following types are defined by Johnson (1961, p. 3-4):

1. Interjections of sounds, syllables, words or phrases: Extraneous sounds or words inserted into the otherwise fluent flow of speech. Examples include *well*, *hmm* and *uh*.

2. Part-word repetitions: A part of a word, namely a phoneme or syllable, is repeated at least twice in a row. An example is *buh-boy*.

3. Word repetitions: A whole word is repeated at least twice in a row. An example is *I was-was going*.

4. Phrase repetitions: Two or more words are repeated in a row. An example is *I was-I was going*.

5. Revisions: The speaker corrects themself, either modifying the semantic content, phonological encoding, or grammatical features of previously uttered speech. An example is *I was-I am going*.

6. Incomplete phrases: A phrase that is abandoned and left unfinished. An example is *She was-and after she got there he came*.

7. Broken words: A word that is not completely pronounced or in which the expected rhythm is broken. An example is *I was g- (pause) -oing home*.

8. Prolonged sounds: Sounds that are longer than normal.

Whereas Duchin & Mysak (1987), Leeper & Culatta (1995) and Searl et al. (2002) all consider the first seven disfluencies defined by Johnson (1961), only Searl et al. (2002) analyse prolongations as well. Moreover, all three studies also analyse tense pauses, contradictorily defined as *silent prolongations or blocks* (Searl et al., 2002, p. 386) or as *barely audible manifestation of heavy breathing or muscular tightening* (Leeper & Culatta, 1995, p. 6). Additionally, Duchin, & Mysak (1987) and Leeper & Culatta (1995) also analyse dysrhythmic phonation, defined as a *disturbance of the normal rhythm or flow of speech within a word* (Leeper & Culatta, 1995, p. 6). Additionally, Leeper & Culatta (1995) also

analyse hesitations, defined as *nontense pauses occurring in the normal forward flow of speech* (Leeper & Culatta, 1995, p. 6).

While more modern studies have more diverse categorisations, many are still reminiscent of Johnson's (1961) categorisations, analysing repetitions, revisions, and interjections/filled pauses (Bortfeld et al., 2001; Keszler & Bóna, 2019; Samani et al., 2017; Weathersby, 2016), prolongations (Keszler & Bóna, 2019; Samani et al., 2017), incomplete phrases (Samani et al., 2017; Weathersby, 2016) and broken words (Weathersby, 2016). Although, in stark contrast to Johnsonian studies, modern studies often add additional disfluency types. While Samani et al. (2017) analyse the undefined *blocks*, the most notable difference compared to Johnson's (1961) categorisation is the treatment of unfilled pauses. Even though most non-Johnsonian studies omit unfilled pauses altogether, those who do analyse them have varying definitions. While Keszler & Bóna (2019) analyse only within word pauses of unspecified length, Samani et al. (2017) and Weathersby (2016) analyse between word pauses as well. Yet, while Samani et al. (2017) define an unfilled pause as being 200 milliseconds or longer, Weathersby (2016) defines an unfilled pause as being 1,000 milliseconds or longer. Note that pauses for audible breathing are counted as disfluencies in the three studies.

Other studies have diverged more from traditional categorisations. For instance, Horton et al. (2010) analyse only filled pauses, but do distinguish between *eh* and *uhm*, whereas Moscoso del Prado Martín (2016) does not clarify which types of disfluencies he analyses, defining them simply as "*rephrasings, filled pauses, etc.*" (p. 954), and calculates only an overall rate of disfluencies without distinguishing between types.

2.1.3. Speech Errors

There is a plethora of different speech errors, such as malapropisms (Levelt, 1989, p. 355) and omissions (Levelt, 1989, p. 396), all of which reveal aspects of the process of lexical selection (Levelt, 1989, p. 234). However, this thesis only treats substitutions. These arise when the target word the speaker intends to produce is substituted for a different word (Levelt, 1989, p. 220). For example, when a speaker intends to warn a listener not to burn their fingers on something hot, they might substitute *fingers* for the word *toes*, producing *Don't burn your toes* (Fromkin, 1973, referred to in Levelt, 1989, p. 218). This example is an associative intrusion, meaning the concept for *fingers* is semantically associated both with the

correct lemma as well as the lemma for *toes* and, for some reason, the incorrect lemma is chosen and the substitution is produced. This type of substitution is the most commonly found, and associative intrusions are often antonyms (i.e. opposites) or co-hyponyms (i.e. words from the same semantic field (Levelt, 1989, p. 219) such as *cow* and *sheep*). Conceptual intrusions are less commonly found since the produced utterance is still typically appropriate in the context and therefore not noticed. These arise when the target word is conceptually related to a different word, for example hypernyms such as *dog* and *animal* (Levelt, 1989, p. 220). Moreover, substitutions can also arise due to intrusions of a recently activated lemma. In the example *The branch fell on the tree*, the word *branch* is associated with the lemma for *tree*, which then substitutes the target word *roof*. Similarly, conceptually planning ahead for the next utterance can also cause a substitution, such as substituting the target word with a word the speaker intended to produce in the next sentence (Levelt, 1989, p. 221).

2.1.4. Speech Rate

While speech rate has seemingly been studied less than disfluencies and speech errors in relation to the speech production process, it is still a highly relevant variable. The speech rate is determined during phonological encoding (Levelt, 1989, p. 306), and is mostly varied by the rate of pauses, with faster speech having fewer pauses (Levelt, 1989, p. 385). Consequently, speech rate is also related to processing, with faster speech being more effortful for the speech production system to produce (Levelt, 1989, p. 487). However, slower speech can actually result in more errors, since it leaves more time for activation to spread to an incorrect lemma (Levelt, 1989, p. 355-359). In contrast, faster speech is associated with an increased violation of rules, such as the main interruption rule (Levelt, 1989, p. 487). Some studies have also suggested that slower speech might result in more disfluencies (Horton et al., 2010; Leeper & Culatta, 1995; Samani et al., 2017), yet other studies find no such relation (Searl et al., 2002). Furthermore, speech rate is also communicative, for instance as speakers tend to speak more quickly towards the end of a sentence to hold the floor and stop their interlocutor from interrupting (Levelt, 1989, p. 306).

While the general term *speech rate* has been used up until this point, there are in fact two main methods of calculating the rate of speech. The first is *speaking rate* (Kendall, 2013, p. 58), also known as the *general rate of production*. The second is *articulation rate* (Kendall, 2013, p. 58). The general rate of production is a measure of speech rate that includes all

pauses, both unfilled and filled, in the calculated measure, whereas the articulation rate excludes all unfilled and filled pauses longer than 200 milliseconds (Kendall, 2013, p. 58-59). The two are not always perfectly correlated (Kendall, 2013, p. 58) and thus provide two different perspectives. The units measured in speech rate also vary, as one can measure either words, syllables or phonemes per minute or second. While all three are used, Kendall (2013, p. 60) claims that measuring syllables is the most accurate. However, there are three main methods for counting syllables, namely by relying on the acoustic signals of speech, on impressionistic counting of sounds, or on the written orthographic representation of speech and syllables. Orthographic counting is the most common method in modern research (Kendall, 2013, p. 62).

2.1.5. Task Complexity

As mentioned, task complexity affects the speech production process. During conceptualisation, the speaker must extensively search their memory to retrieve information relevant for planning the utterance. The processes of searching memory and planning are under executive control and therefore require the speaker's attention (Levelt, 1989, p. 126). Consequently, complex tasks, which require more attention, result in more problems in the speech production process. For instance, Bortfeld et al. (2001) found that speakers produce more disfluencies when describing complex abstract shapes than pictures of children, Goldman-Eisler (1968, referred to in Levelt, 1989, p. 128) found that speakers produced more disfluencies when interpreting the message of a cartoon than when simply describing it, Good & Butterworth (1980, referred to in Levelt, 1989, p. 128) found that speakers produced more disfluencies when describing unfamiliar routes than familiar routes, and Goldman-Eisler (1958, referred to in Kendall, 2013, p. 21) found that pause rates increase with task complexity. In sum, the greater the complexity of the task, the more issues and disfluencies can be expected to arise during speech.

2.2. Disfluencies and Age

Several studies have investigated the correlation between disfluencies and age throughout the adult lifespan. First of all, Duchin & Mysak (1987), Leeper & Culatta (1995) and Searl et al. (2002) rely on the discussed disfluency types defined by Johnson (1961). However, while both Duchin & Mysak (1987) and Leeper & Culatta (1995) study English spontaneous speech, semi-spontaneous picture descriptions and oral reading of participants between circa

20 and circa 85 years old, Searl et al. (2002) study the English spontaneous speech of seven speakers between the ages of 100 and 103.

As for their results, both Duchin & Mysak (1987) and Searl et al. (2002) find no significant increase in disfluency rates in old age, with the former finding a disfluency rate, calculated per 100 words regardless of speaking situation, ranging from 7.37% (i.e. 7.37 disfluencies occur per 100 words) for speakers between 21 to 34 years old to 7.83% for speakers older than 75, and the latter finding a normal rate of 6.2% for the centenarians. In contrast, Leeper & Culatta (1995) find no increase in disfluency rates for semi-spontaneous and spontaneous speech, yet note a significant increase with age in reading aloud. They postulate that this increase in disfluencies might be caused by decreased visual acuity, or due to the fact that older adults might struggle with the more rigid semantic, syntactic and lexical structure of the written passage, preferring to freely structure their own speech.

Second, for studies not explicitly relying on Johnson's (1961) disfluency types, the results are varied. Studies that find a positive correlation between disfluency rates and age are Bortfeld et al. (2001), in their study of English director/matcher picture descriptions of speakers between 20 and 70 years of age, Horton et al. (2010), in their corpus study of English prompted conversations of speakers between 20 and 67 years of age, and Samani et al. (2017), in their study of prompted Persian monologues of speakers between 20 and 90 year of age. Bortfeld et al. (2001) find that the average disfluency rate for their cohort with an average age of 28 is 5.55% compared to 6.65% in their oldest age group with an average age of 68, while Samani et al. (2017) find that the average disfluency rate, excluding unfilled pauses, for adults aged between 20 to 44 is 11.07%, compared to 16.97% in adults aged 75 to 90. Both Bortfeld et al. (2001) and Samani et al. (2017) find that all types of disfluencies increase in old age, but with no notable difference between young and middle-aged speakers, who produce disfluency rates of 5.69% (Bortfeld et al., 2001) and 11.24% (Samani et al., 2017). These increases are explained by Samani et al. (2017) to stem from decreased articulatory control and cognitive processing speed which results in issues in lexical retrieval in old age. Moreover, while Bortfeld et al. (2001) find that both *eh* and *uhm* increase with age, Horton et al. (2010) find that although overall rates of filled pauses increase in old age, the rate of *uhm* decreases. Horton et al. (2010) claim that since *uhms* are associated with more holistic message-level planning of the utterance in its entirety, planning strategies must change with age.

In contrast, Keszler & Bóna (2019) find no significant increase in disfluency rates with age in their longitudinal study analysing the spontaneous speech of seven Hungarian speakers from when they were middle-aged to when they were older than 75 years old. They found that while certain disfluencies become more common with age for certain speakers, other speakers spoke at an equally fluent level both when they were middle-aged and older than 75. Consequently, Keszler & Bóna (2019) claim that disfluency rates vary on an individual level, and that old age must not entail less fluent speech. The most salient increase in disfluency rates in relation to old age found by Keszler & Bóna (2019) are in pause frequency and pause duration. Yet, as mentioned, since Keszler & Bóna (2019) include physiological pauses for breathing in their definition of unfilled pauses, they theorise that these increases might be caused by physiological changes in the breathing apparatus, since age is positively correlated with more frequent and longer pauses for breathing (Bóna, 2018; Gyarmathy, 2019, referred to in Keszler & Bóna, 2019).

Moreover, other modern studies have yielded mixed results. Similar to Leeper & Culatta (1995), Weathersby (2016), in her study of 64 to 90 year old participants' English spontaneous speech, picture descriptions and oral reading, finds that disfluency rates only vary with age in reading, with older participants producing more disfluencies. Again similar to Leeper & Culatta (1995), Weathersby (2016) attributes this increase to decreased visual acuity in relation to age. Moreover, Moscoso del Prado Martín (2016), relying on the same corpus as Horton et al. (2010), analyses prompted English conversations of 20 to 68 year olds, calculating the number of disfluencies per clause. Interestingly, he concludes that disfluency rates decrease linearly for women throughout their lifespan, with older women producing fewer disfluencies than younger women. However, for men, he finds a linear decrease until about the age of 45, after which disfluency rates increase linearly. Moscoso del Prado Martín (2016) does not explain why disfluency rates might vary with age.

Lastly, table 1 is a summary of the disfluency rates found in different studies. The averages are across all speaking situations.

Table 1: Disfluency rates in spontaneous and/or semi-spontaneous speech found across studies. The age groups shown are the cohorts created by the respective authors. The rates are calculated as disfluencies per 100 words, except for in Keszler & Bóna (2019) and Weathersby (2016), where they are calculated as disfluencies per 100 syllables.

Study	Disfluency rate
Bortfeld et al. (2001)	Total average: 5.97%. Average age 28: 5.55%[a] Average age 48: 5.69% Average age 68: 6.65%
Duchin & Mysak (1987)	Total average: 6.83%[b] Age 21-30: 6.99% Age 45-64: 7.62% Age 55-64: 7.08% Age 65-74: 5.64% Age 75-91: 6.83%
Leeper & Culatta (1995)	Total average: 4.87% Age 25-34: 6.36% Age 55-64: 5.2% Age 65-74: 4.23% Age 75-84: 3.95% Age 85+: 5.23%
Samani et al. (2017)	Total average: 12.9%[b, c] Age 20-44: 11.07% Age 45-59: 11.24% Age 60-74: 12.32% Age 75-90: 16.97%
Searl et al. (2002)	Age 100-103: 6.2%
Keszler & Bóna (2019)	Age 49-90: 0-7%[d]
Weathersby (2016)	Age 64-91: 7.8%

[a] *Bortfeld et al. (2001) provide only the average of each age group instead of accounting for the age range of their participants.*
[b] *The study does not provide an overall average, which has instead been calculated by the author of the current study. Consequently, the varying sizes of cohorts are not accounted for in this average, as it is simply the average of the rates per cohort in total.*
[c] *Samani et al. (2017) exclude unfilled pauses from their disfluency rates.*
[d] *Keszler & Bóna (2019) provide no average disfluency rates, accounting only for the range of disfluencies between participants.*

2.3. Speech Errors and Age

Few studies seem to have been conducted on the relation between speech errors and aging. Importantly, previous studies rely on strictly controlled stimuli, namely reading.

For instance, Gollan & Goldrick (2019) allow two groups of monolingual speakers of English, a younger group with a mean age of 20 and an older group with a mean age of 72, to

read different types of paragraphs of text. Each participant reads normal texts, texts where nouns are randomly exchanged across positions, and texts where adjacent words are exchanged. Looking at several error types, one of which is substitutions, Gollan & Goldrick (2019) find that the rate of error does not vary significantly with age on manipulated paragraphs, but that older adults produce significantly more errors when reading normal paragraphs. Additionally, older adults self-correct their errors less often than younger adults. Gollan & Goldrick (2019) conclude that aging imposes limitations on the monitoring of speech and leads the cognitive speech production process to be more error prone.

Moreover, MacKay & James (2004) show young adults with an average age of 19 and older adults with an average age of 72 written words and ask them to produce certain written letters as different sounds than they typically correspond to. For instance, if the prompted word is *pill*, the participant should say *bill* (Mackay & James, 2004, p. 97). If the participant does not correctly transform the word prompt into the desired outcome, it is counted as an error. Looking at several types of speech errors, including substitutions, MacKay & James (2004) find that overall error rates increase with age, whereas substitutions are more common in younger participants. As for self-corrections, the results are mixed and therefore insignificant. MacKay & James (2004) argue that older adults are more likely to produce speech errors due to cognitive changes in the speech production process, mostly due to a decline in the cognitive transmission speed.

Contrastingly, Vousden & Maylor (2006) find no correlation between speech errors and age. Young participants with an average age of 21, and older participants with an average age of 72, are made to read tongue twisters, such as *five frantic fat frogs* (p. 48), aloud. Participants are instructed to match their speech rate to a beat set by a metronome, producing one word per beat, to test for differences at a slow and fast pace respectively. Analysing a variety of error types, including substitutions, Vousden & Maylor (2006) find that older adults are incapable of maintaining the fast pace of speech, but do not produce significantly more errors at the slow pace.

2.4. Speech Rate and Age

Since the two are interconnected, most studies of fluency in relation to aging analyse speech rate as well. This section summarises and compares previous methods and findings of speech rate in relation to aging.

Firstly, studies vary in how they measure speech rate. Most studies rely on measuring either syllables per minute or words per minute, meaning the total number of syllables or words spoken divided by total speaking time including pause time (Duchin & Mysak, 1987; Horton et al., 2010; Leeper & Culatta, 1995; Samani et al., 2017; Searl, et al., 2002). This is equivalent to the general rate of production. All aforementioned studies find at least partial evidence that older speakers produce slower speech. As for words per minute, the oldest age group speaks at a slower pace, ranging from, on average, 113 words per minute (Searl et al., 2002) to 169 words per minute (Horton et al., 2010), compared to the youngest age group, ranging from, on average, 185 words per minute (Duchin & Mysak, 1987) to 194 words per minute (Horton et al., 2010). As for syllables per minute, all studies again find that the oldest age group speaks at a slower pace, ranging from, on average, 125 syllables per minute (Samani et al., 2017) to circa 200 syllables per minute (Leeper & Culatta, 1995), compared to the youngest age group, ranging from, on average, circa 240 syllables per minute (Duchin & Mysak, 1987) to 308 syllables per minute (Samani et al., 2017).

Samani et al. (2017) propose several reasons for why speech rate decreases with age. They argue that older adults might monitor their speech more closely than younger participants, meaning they are more careful and thus speak more slowly. They also propose that decreased speech rates might be caused by decreased cognitive abilities in old age, or by decreased muscular strength and control of the articulators.

Secondly, while Duchin & Mysak (1987) find a decrease in the general rate of production in older participants across all speaking situations, Leeper & Culatta (1995) find no decrease in monologues or picture descriptions, instead noting only a negative correlation between age and speech rate in oral reading, decreasing from a general production rate of circa 250 syllables per second for 25 to 34 year olds to circa 200 syllables per second for those older than 85. The study suggests that the decrease likely stems from the visuospatial demands of reading, since tasks with a major visuospatial component induce more negative performance

in older speakers as compared to young speakers (Knoefel, 1990, referred to in Leeper & Culatta, 1995).

Thirdly, only Samani et al. (2017) calculate the articulation rate of their participants. They find a significant decrease with age, with speech ranging from an average of 211 syllables per minute for older adults compared to an average of 323 syllables per minute for young adults.

Fourthly, while all aforementioned studies find at least a partial decrease in speech rate in older speakers, there is no consensus on whether this decrease is linear with age or not. Duchin & Mysak (1987) claim that participants in their twenties speak significantly faster than middle-aged participants, who in turn speak significantly faster than participants aged 75 to 91. Similarly, Leeper & Culatta (1995) argue that speech rate decreases linearly with age, decreasing by about 10 syllables per minute per cohort. In contrast, Samani et al. (2017) find that young and middle-aged participants speak at comparable levels whereas participants aged 60 to 90 speak significantly slower.

Table 2 summarises the speech rates found in the studies discussed above.

Table 2: The average words per minute (wpm) and syllables per minute (spm) found across the studies. All measures are stated as the general rate of production (GPR) except for in Samani et al. (2017), where they are also stated as articulation rate (AR).

Study	Speech rate
Duchin & Mysak (1987)	Total average[a]: 161 wpm, 210 spm Age 21-30: 185 wpm, 240 spm Age 45-64: 157 wpm, 203 spm Age 55-64: 167 wpm, 215 spm Age 65-74: 156 wpm, 201 spm Age 75-91: 140 wpm, 189 spm
Horton et al. (2010)	Average age 20: 194 wpm Average age 60: 169 wpm
Leeper & Culatta (1995)	Total average: 225 spm[a, b] Age 25-34: Ca 250 spm Age 55-64: Ca 238 spm Age 65-74: Ca 226 spm Age 75-84: Ca 213spm Age 85+: Ca 200 spm
Samani et al. (2017)	Total average:

	Age 20-44: 308 spm (GPR), 323 spm (AR)
	Age 45-59: 312 spm (GPR), 301 spm (AR)[c]
	Age 60-74: 199 spm (GPR), 299 spm (AR)
	Age 75-90: 125 spm (GPR), 211 spm (AR)
Searl et al. (2002)	Age 100-103: 113 wpm

[a] *The study does not provide an overall average, which has instead been calculated by the author of the current study. Consequently, the varying sizes of cohorts are not accounted for in this average, as it is simply the average of the rates of each cohort in total.*
[b] *The study provides their data only in visual graphs, meaning no exact numbers are available. The averages presented here are my own visual estimations.*
[c] *Samani et al. (2017) claim that the articulation rate for this cohort is lower than the general production rate, which should not be possible. This is likely an error.*

2.5. Fluency and Other Factors

While this thesis concerns only age in relation to fluency, other sociolinguistic factors might influence fluency. For instance, many authors propose that males are more disfluent than females, both in diagnosed stutterers (Yairi et al., 1996, referred to in Drayna et al., 1999) and in healthy speech (Bortfeld et al., 2001; Leeper & Culatta, 1995; Moscoso del Prado Martín, 2016; Weathersby, 2016). Similarly, sex is also argued to be correlated with speech rate, with males speaking faster than females in a variety of languages (Byrd, 1992; Byrd, 1994; Fitzsimons et al., 2001; Quené, 2005; Quené, 2008; Verhoeven et al., 2004; Whiteside et al., 1986; Yuan et al., 2006, referred to in Van Borsel & De Maesschalck, 2008). Furthermore, speech rate varies both depending on the language of the speaker (Amino & Osanai, 2015; Coupé et al., 2019; Hilton et al., 2011) and on the specific accent of the speaker within a language (Byrd, 1994, referred to in Kendall, 2013, p. 26; Jacewicz et al., 2009; Verhoeven et al., 2004). Additionally, the level of education is theorised to affect disfluency rates, yet Weathersby (2016) finds no correlation between fluency and education.

2.6. The Current Study

The results of previous studies regarding the correlation between age and fluency are inconclusive. Moreover, some variables have been studied markedly less than others in relation to age. Most importantly, unfilled pauses have been neglected. As far as the author is aware, only Samani et al. (2017) analyse unfilled pauses longer than 200 milliseconds both within and between words in relation to age. However, Samani et al. (2017) do not exclude physiological pauses for breathing, which are likely to skew the results by making older adults appear more disfluent. Additionally, while task complexity has been documented to

influence fluency, it is not known whether task complexity influences the fluency of different age groups to different degrees.

The current study is an experimental cross-sectional cohort study designed to compare the frequencies and types of disfluencies, speech errors, speech rate and pause duration in six younger, six six middle-aged, and six older speakers of Swedish. Importantly, unfilled pauses, excluding physiological pauses, are analysed. Speakers are tested on both a simple and a complex task to investigate the effect of task complexity. Below is the overarching research question the study is centered around as well as the more specific questions the study aims to answer.

- Does fluency change throughout the lifespan?
  - Does fluency differ between age groups?
    1. Is there a difference in the overall rate of disfluencies across age groups?
    2. Is there a difference in the rates of different types of disfluencies across age groups?
    3. Is there a difference in pause duration across age groups?
    4. Is there a difference in the overall rate of speech errors across age groups?
    5. Is there a difference in speech rate across age groups?
  - Does task complexity influence fluency differently across age groups?
    6. Does the overall rate of disfluencies differ with task complexity across age groups?
    7. Do the rates of different types of disfluencies differ with task complexity across age groups?
    8. Does pause duration differ with task complexity across age groups?
    9. Does speech rate differ with task complexity across age groups?

3. Method

The current study is an experimental cohort design utilising two levels of semi-structured stimuli to elicit data. The correlation between the independent variables age as well as

stimulus complexity and the dependent variables rate of disfluencies, speech errors, and speech rate are studied.

3.1. Participants

All participants in the study were native, monolingual speakers of Swedish. As shown, different languages have different speech rates. No bilinguals were chosen since the influence of a second native language might affect speech rate or fluency. All participants were divided into three cohorts, namely young (Y), middle-aged (M) and old (O) and given a number between 1 and 6. Young participants have an average age of 20 and range from 19 to 20 years old (*n*=6), middle-aged participants have an average age of 52 and range from 50 to 56 years old (*n*=6), and old speakers have an average age of 80 and range from 71 to 85 years old (*n*=5). A middle-aged group is important to determine if differences in fluency are linear throughout the lifespan. Furthermore, as discussed, other sociolinguistic factors than age might influence fluency. Therefore, each participant was questioned regarding their sex, self-perceived accent, region where they grew up, and their highest level of completed education (cf. Appendix A). Both accent and region of origin are important since some participants provide mixed answers regarding accents, such as O3 who said they speak a mix of *sörmländska* (accent from the region of *Södermanland*) and *skånska* (accent from the region of Scania). Since O3 stated that they grew up in *Södermanland*, *sörmländska* was considered to be their accent. While this method is not perfect, it avoids multiple values for accents which would unnecessarily complicate this study since it is not primarily interested in dialects. Moreover, if a participant stated that they had a local accent, such as Y4 who stated that they spoke *lundensiska* (the accent associated with the city and surrounding area of Lund, Scania), their accent was generalised as *skånska*. Note that since the study was conducted in Scania, most participants are speakers of Scanian. While no analyses of sex, education or accent are performed due to the small sample size and a lack of time, they are important for comparability with future studies.

Most participants were sourced based on opportunity, that is, relying on those who are available. Many participants were sourced from the author's acquaintances (Y1-Y6; M1-2), some were sourced by word of mouth from acquaintances (M3-M6; O1-5) and one was sourced through flyers (O6). However, participant O6 withdrew their consent from the study. Consequently, the young and middle-aged cohorts consist of six participants each and the

older cohort consists of five participants, resulting in a total sample of 17 participants. While the sample is small, data elicitation and transcription are time consuming processes. With the limited time and resources available for the study, a sample of 17 participants is deemed an acceptable size that is still large enough to draw some conclusions. However, it should be noted that the generalisability of the findings and the representativeness of the data is limited due to the small and unrepresentative sample.

Table 3: Participant demographics.

Participant	Sex	Age	Degree	Accent	Region
Y1	Female	20	High School	Skånska	Skåne
Y2	Female	20	High School	Skånska	Skåne
Y3	Female	20	High School	Skånska	Skåne
Y4	Male	19	High School	Skånska	Skåne
Y5	Male	20	High School	Skånska	Skåne
Y6	Male	20	High School	Skånska	Skåne
M1	Female	51	High School	Skånska	Skåne
M2	Female	50	Bachelor's	Västgötska	Västra Götaland
M3	Female	52	Vocational	Skånska	Skåne
M4	Male	53	PhD	Skånska	Skåne
M5	Male	56	Bachelor's	Skånska	Skåne
M6	Male	52	Master's	Skånska	Skåne
O1	Female	85	High School	Sörmländska	Södermanland
O2	Female	85	Bachelor's	Skånska	Skåne
O3	Female	71	Bachelor's	Sörmländska	Södermanland
O4	Male	84	PhD	Skånska	Skåne
O5	Male	73	High School	Skånska	Skåne

3.2. Materials and task

As discussed, the complexity of the task affects fluency, with more complex tasks resulting in higher rates of disfluency. Moreover, while spontaneous speech has the benefit of being more naturalistic than more controlled elicitations, it is often difficult to anticipate what the participant intends to say (Kavé & Goral, 2017). With more controlled stimuli, the experimenter has a better understanding of what the speaker intends to produce, and repairs and errors are therefore more easily understood (Levelt, 1989, p. 75). Relying on these

principles, the current study utilised semi-spontaneous stimuli derived from Melinger & Kita (2007).

The stimuli used by Melinger & Kita (2007) are networks, similar to maps, with coloured nodes. Importantly, the stimuli consists of two levels, one simple and one complex. This allows the current study to analyse whether or not task complexity affects the speech production of different cohorts differently. At the simple level, the participant is allowed to look at and memorise the stimulus for 15 seconds after which the stimulus is removed and the participant has to describe, from a fixed starting point indicated by an arrow, which colour the first node is, which direction you have to go to reach the next node, and so on until the end of the network. At the complex level, the participant is presented with a different stimulus for 15 seconds after which they describe it in the same manner as the simple stimuli. However, two additional colours, namely a green and an orange node, are added which act as secondary tasks to complicate the utterance. When the participant reaches the green node, they must explain the location of the green node in relation to the starting point. Afterwards, the participant continues describing the network from the green node. When the participant reaches the orange node, they must retell which letter, written as a capital letter in bold black, is inside the node as well as which letter follows in the alphabet. Afterwards, the participant continues describing the network from the orange node. Each complex stimulus contains one green and one orange node. At both levels respectively, three stimuli have seven nodes and two stimuli have six nodes.

In Melinger & Kita's (2007) study, simple networks can be either deterministic, meaning there is only one possible route the nodes follow, or non-deterministic, meaning the nodes diverge into several possible routes, whereas all complex stimuli are deterministic. For the current study, only deterministic networks were used since a third level of complexity was deemed superfluous since the research question regarding complexity will be sufficiently answered by only two levels. Moreover, Melinger & Kita (2007) rely on a director-matcher system, in which a participant describes the stimulus to another participant who answers written questions regarding the relevant stimulus. In the current study, a director-matcher task was not used, primarily since it would be infeasible with opportunity sourcing to locate two participants, or a confederate, per elication, and since a dialogue would theoretically result in more communicative hesitations and speech rate modifications which may skew the results. However, the absence of a director-matcher system might entail that participants were more

aware of the experimental nature of the elicitation, consequently rendering their speech less naturalistic.

Note that since Melinger & Kita (2007) do not provide more than a few examples of their stimuli, original recreations based on their designs were created for the current study (Appendix B). Each stimulus was labeled with S for Simple or C for complex and given a number between 1 and 5. Figure 1 shows stimulus C5 and its description provided by Y3, with disfluencies removed to increase readability.
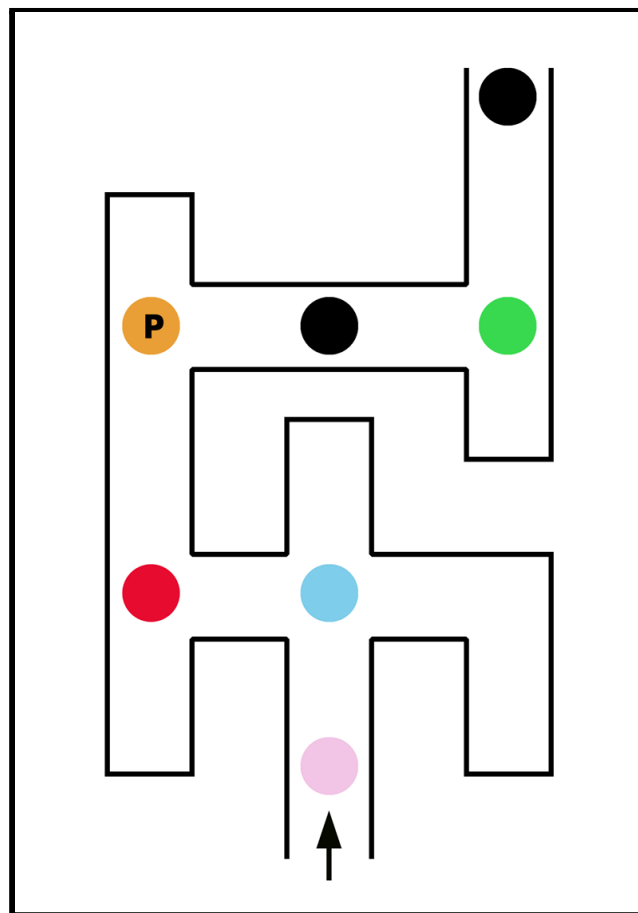

Figure 1: Complex stimulus five (C5).

Y3: man går upp till en rosa sen går man upp till en blå sen går man vänster till en röd sen går man upp till en orange där det står P Q är nästa sen går man höger till en svart sen går man höger till en grön som är snett upp nästan rakt från där man börjar men lite snett till höger och sen går man upp till en svart

Y3: you go up to a pink then you go up to a blue then you go left to a red then you go up to an orange where it says P Q is next then you go right to a black then you go right to a green which is diagonally

up almost straight from where you start but a little diagonally to the right and then you go up to a black

3.3. Procedure

Each elicitation was performed in a quiet room. Relying on opportunity sourcing, the study adapted to the needs and wishes of the participants, meaning elicitations were performed in different locations depending on the participant's preference. Some were conducted in the home of the experimenter (Y1-2; M2; M6), some in locations at university campus (Y4-6; M5), and some in the home of the participant or a relative of the participant (Y3, M1; M3-4; O1-5). All locations were ensured to be sufficiently quiet. Elicitations were also performed at varying times of day based on the participant's availability and choice. While more uniform procedures for elicitation would ensure less possible discrepancies in the results, stricter rules for participation would undoubtedly have resulted in a smaller sample, especially since many older participants were understandably unwilling to leave their own residence during an ongoing pandemic. Furthermore, each elicitation is recorded with the experimenter's phone. While a proper microphone would have been advantageous, a phone was the best recording device available for the study.

During each elicitation, the experimenter adhered to a script (Appendix A) to ensure that each participant received the same instructions. This limits discrepancies in results. Each elicitation commenced by providing the participant with two copies of a printed consent form, one for the participant and one for the experimenter. The consent form informed them of the basic purpose and outline of the study, their anonymity, their right to terminate the study at any time, how their data will be stored and that they can request to have it deleted at any time, and contact information for both the experimenter and the supervisor (Appendix C). Next, the participant was interviewed for background information (Appendix A), after which the participant was informed of the task instructions for the simple material. No verbal examples were given so as not to influence their speech. The participant was encouraged to ask questions if they did not understand. The participant was also informed that the experimenter will sit quietly and only look at the stimulus during the description in order to make the speaking situation as close to a monologue as possible. The purpose of this was to reduce communicative pauses and modifications of speech rate. Subsequently, an unrecorded practice round of the simple material was performed. The stimulus was presented using Microsoft PowerPoint timed to display the stimulus for 15 seconds. If the participant did not

describe the stimulus as requested, for instance leaving out directions between nodes, they were provided with verbal examples of how to describe the material. Thereafter, the participant was informed of the task instructions for the complex stimuli and a practice round of the complex material ensued. After the practice round, if misinterpretations of the instructions occured, the participant was provided with verbal examples. Next, the participant was presented with the real stimuli, consisting of four simple and four complex stimuli, one at a time. These elicitations were audio-recorded. While all participants had the same stimuli as practice rounds, namely S1 and C1 in that order, during the actual recorded experiment the order of the simple and complex stimuli was counterbalanced to limit the influence of fatigue. Lastly, audio-recording stopped and the participant was debriefed and allowed to ask questions freely. Since debriefing is a dialogue, no script was followed for this part to encourage participants to interrupt and ask questions freely.

4. Data Treatment

4.1. Relevant Data

The data relevant for this essay are data connected to the speech production process. As for disfluencies, the current study relies on the disfluency types described by Graziano & Gullberg (2018), namely filled pauses, unfilled pauses, interruptions, prolongations, and repetitions. First of all, these disfluencies are adequately detailed, meaning that they capture most major types of disfluencies without delving into excessive subtypes. Second, unlike many other studies, Graziano & Gullberg (2018) provide clear definitions and examples of their disfluency types, rendering them easy to reproduce. However, certain modifications to their disfluency types are made for the current study. Most importantly, Graziano & Gullberg (2018) only analyse intra-clausal disfluencies (i.e. disfluencies occuring within a clause). In the current study, such a distinction is not possible. Due to the nature of the task, several participants speak in an unnatural manner akin to telegraph speech that does not correspond to traditional clauses. For instance, participant Y2 produces only five verbs in a total of 145 words, and five out of eight stimuli were described completely without verbs. Consequently extra-clausal pauses (i.e. disfluencies occurring between clauses), which are not true disfluencies but simply pauses for planning, are included in the disfluency rates. Additionally, the current study is concerned with the duration of pauses which are consequently annotated in the transcriptions. Moreover, Graziano & Gullberg (2018) include a *combination* category where several errors of different types occur in sequence. The current study excludes this

category since it is more concerned with the quantitative rate of overall disfluencies than a more qualitative analysis of specific combinations. While this would certainly be interesting, it is simply deemed as outside the scope of this paper and is not possible within the timeframe.

As for speech errors, these are rarer and more unpredictable than disfluencies, occurring as seldom as once per 1,000 words (Garnham et al., 1982, referred to in Levelt, 1989, p. 199). As a result, speech errors are coded post-hoc based on what appears in the data.

Regarding speech rate, both the general rate of production and the articulation rate are calculated. Moreover, the number of words, syllables and utterance length in milliseconds are analysed to provide relative values of fluency. Since the study is also concerned with task complexity, each of the aforementioned values are calculated per stimuli and generalised across simple and complex stimuli elicitations respectively.

4.2. Transcription and Coding

The transcription and coding of the utterances is a significant part of the study. Firstly, transcription is performed with a sort of allegro form, meaning non-pronounced sounds are omitted (Litosseliti, 2018, p. 92). Specifically, only non-pronounced syllables are omitted since these affect the speech rate values, and relying on orthography would skew the results. However, the reduction of singular phonemes but not complete syllables are not noted since they are not relevant to the study. Inaudible or untranscribable segments are transcribed in capital letters within brackets, such as *[WHISPER]* or *[LAUGHTER]*. Some participants produced speech between stimuli, that is, in the 15 seconds where the participant is memorising, such as in (1) produced by M1 when describing stimulus C4. Speech produced between stimuli is not influenced by task complexity, and therefore likely more fluent than the actual elicitations, meaning that including it in the analysis would skew the results.

    (1) *åh vad fint att man gick ner för en gångs skull*

        oh how nice that you went down for once

Secondly, disfluencies are coded according to the modified definitions of Graziano & Gullberg (2008, p. 5) provided below.

- **Filled pauses**: *eh* and *uhm*, each coded separately and immediately followed by the duration of the filled pause written as (X,), where the X stands for the duration in milliseconds. The duration is measured using the program PRAAT. Note that there is no minimal duration for a filled pause (Rose, 2017). For example: [M2: "...*vänster rakt upp från eh(214,) den första...*" "...left straight up from eh(214,) the first…"]

- **Unfilled pauses**: A silence exceeding 200 milliseconds that is not at the end or start of a speech turn, and is not caused by audible breathing. Transcribed as "*(X.)*", where the X stands for the duration of the pause in milliseconds. The duration is measured using PRAAT. For example: [M1: "...*rakt upp till (298.) orange...*" "...straight up to (298.) orange…"]

- **Interruptions**: When the participant interrupts their own speech and starts over. Transcribed with an apostrophe. For example: [Y4: "...*till höger om' till vänster om den rosa...*" "... to the left' to the right of the pink…"]

- **Prolongations**: Also known as lengthenings. When the participant stretches a sound beyond its normal length. Transcribed with a colon after the lengthened sound. For example: [O3: "...*en orange s:om har bokstaven G...*" "...an orange th:at has the letter G…"]

- **Repetitions**: When the participant repeats a phoneme, syllable or word at least twice. In line with previous studies, regardless of the number of times a segment is repeated it is only counted as one repetition. For example: [O5: "...*en grö* grön till vänster...*" "... a gree* green to the left…"]

Thirdly, speech errors are coded. After transcribing, substitutions are deemed to be the only consistently appearing type. These are annotated with the symbol < to increase searchability. While coding for speech errors is always impressionistic, substitutions are found either when participants correct themselves, such as in (2) produced by O5 when describing stimulus S4, or when the substituted word is logically impossible in the circumstance and consequently deemed to be an error of speech instead of memory, such as in (3) produced by Y1 when describing stimulus C4.

(2) *svart till väns'< (337.) höger menar jag*

black to the lef'< (337.) right I mean

(3) *den gröna låg nästan jämte med den översta< pricken men lite snett ovanför*

the green was almost beside the top< dot but a little diagonally above

In (3), the participant knows they should describe the green node in relation to the first node, yet they accidentally describe it in relation to the top node. Obviously, the green cannot be above the top node, since that would entail that the green node is the top one. Additionally, the green node is almost beside yet a little diagonally above the *first* node, not the top node, which suggests that the participant produced a speech error.

Fourthly, the utterance length per elicitation is measured in milliseconds using the program PRAAT. Lastly, the number of words and syllables per elicitation is calculated based on the transcription of phonological syllables. For words, only finished words are counted, meaning that utterances abandoned mid-word, or repeated syllables or phonemes, are not counted. For syllables, all finished syllables, whether or not the word itself is complete, are counted. However, filled pauses, whispers and laughter are counted neither as words nor syllables. Whispers, laughter and similar are not counted since it is impossible to determine the number of words and syllables produced. Filled pauses are not counted under the assumption that they are only disfluencies and not true words.

4.3. Analysis

Disfluencies are calculated as percentages relative to the total number of syllables produced by the participant. The percentages are calculated by dividing the total number of disfluencies by the total number of produced syllables. Rates are calculated in relation to syllables since it is a more accurate measurement not affected by varying language specific word length or phonological reductions. The rates are presented for the simple and complex stimuli elicitations, respectively, as well as the total overall rates regardless of task complexity. Rates are calculated both for each disfluency type and for all disfluencies overall.

Regarding substitutions, no quantitative analysis is possible due to the small number found in the data. Instead, substitutions are qualitatively described by analysing what typical substitutions appear in each cohort. Therefore, the type of intrusion (associative or conceptual), the location of the error (mid or post word), the edit signal used (if any), and the presence or absence of a hesitation following the error, are analysed. These phenomena are clearly visible and consequently easily observable.

Moreover, for speech rate, both the general rate of production and the articulation rate is calculated. Measurements are calculated per syllable. The general rate of production is calculated by dividing the total number of syllables by the total utterance time, while the articulation rate is calculated by dividing the total number of syllables by the total utterance time excluding filled and unfilled pauses. Speech rates are calculated for the simple stimuli elicitations, the complex stimuli elicitations, and overall.

The data is analysed with descriptive and inferential statistics respectively. Descriptive statistics present the means, medians, standard deviations, etc. Several variables are not sufficiently quantitatively large to perform inferential statistics, meaning descriptive statistics must suffice.

Overall disfluency rates, overall speech rate values, and the duration of unfilled and filled pauses respectively are all based on large amounts of data. Consequently, these values can be analysed with inferential statistics. To determine whether age groups differ on the aforementioned variables, non-parametric Kruskal-Wallis tests are performed with follow-up Mann-Whitney tests in case of main effects. Importantly, these tests treat age as a categorical variable, that is cohorts, and not a continuous one. As for task complexity, Wilcoxon signed rank tests are used to examine differences between simple and complex elicitations within each cohort to determine whether task complexity is a relevant factor in the results. Statistical tests are performed using the program Jamovi.

5. Results

In total, the data consists of 85 stimuli descriptions from 17 speakers totaling circa 131 minutes of recordings, which is reduced to circa 65 minutes of speech when removing the 15 second memorising period for each stimulus. As for the relevant data, 6,412 words, 8,322 syllables, 1,012 unfilled pauses (including extra-clausal silences), 157 *eh*, 74 *uhm*, 132 interruptions, 121 prolongations, 27 repetitions and 18 substitutions were collected. The total disfluency count is 1,523.

5.1. Age and Fluency

5.1.1. Disfluencies

The first research question of the study is *Is there a difference in the overall rate of disfluencies across age groups?* Figure 2 presents overall disfluency rates across individuals which reveals considerable individual differences. Descriptively, the young cohort had an average overall rate of disfluencies of 17.4% (*SD* = 6.3%), the middle-aged had an overall rate of 22% (*SD* = 6.4%), and the older had an overall rate of 19.9% (*SD* = 7.7%). However, there was no main effect of group as shown by a Kruskal-Wallis test ($\chi^2(2)$ = 2.25, *p* = 0.324).
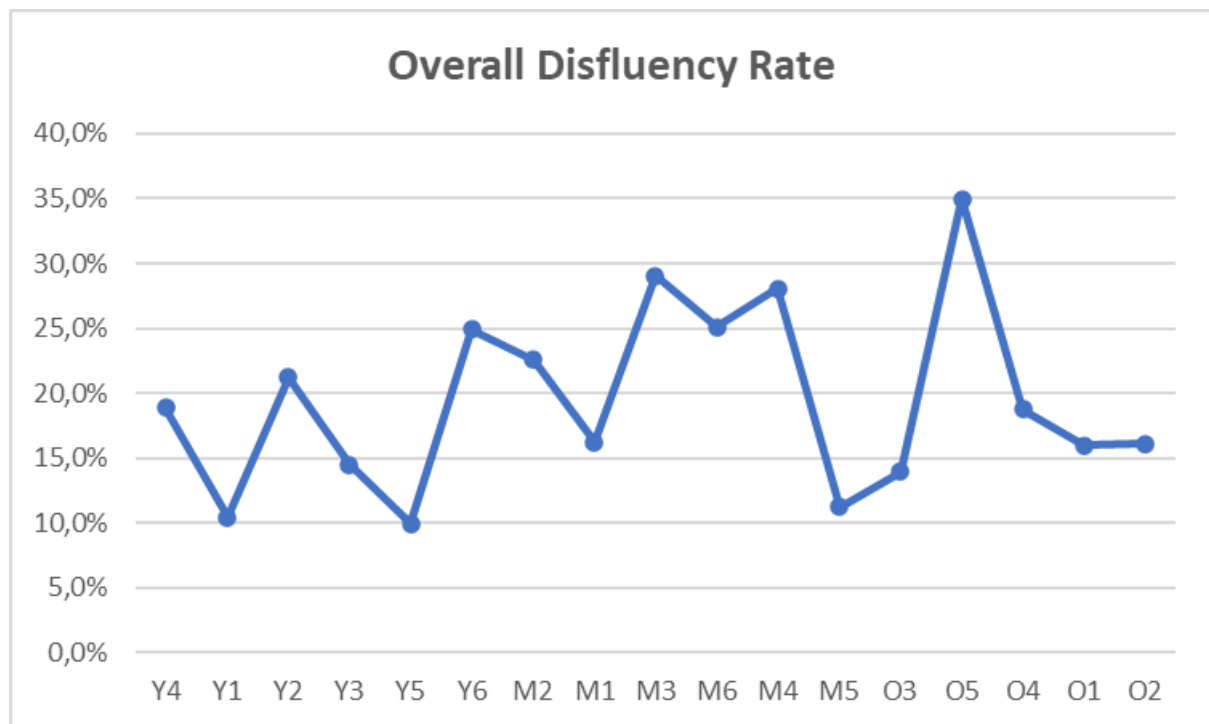
Figure 2: Overall disfluency rates per participant, ordered by age.

The second research question is *Is there a difference in the rates of different types of disfluencies across age groups?* Table 4 presents the rates of disfluency types per cohort. Descriptively, several disfluency types appear to vary with age. As for the rate of interruptions, the young cohort had an average of 1.14% (*SD* = 0.52%), the middle-aged cohort had an average of 1.26% (*SD* = 0.6%), and the older cohort had an average of 2.26% (*SD* = 1.48%), indicating a notable increase with age. Similarly, for the rate of repetitions, the young cohort had an average of 0.02% (*SD* = 0.04%), the middle-aged cohort had an average of 0.21% (*SD* = 0.21%), and the older cohort had an average of 1.06% (*SD* = 0.49%). Contrastingly, for prolongations, the young cohort had an average rate of 1.3% (*SD* =

0.69%), the middle-aged had an average of 1.64% (*SD* = 0.65%), and the oldest had an average of 0.68% (*SD* = 0.65%).

Table 4: Aggregated rates of disfluency type per cohort.

Cohort	Unfilled Pauses	Eh	Uhm	Interruption	Prolongation	Repetition
Young	13.03%	1.74%	0.1%	1.14%	1.3%	0.02%
Middle-aged	15.84%	1.83%	1.29%	1.26%	1.64%	0.21%
Old	13.69%	1.95%	0.42%	2.26%	0.68%	1.06%
Total Average	14.19%	1.84%	0.61%	1.55%	1.21%	0.43%

The third research question is *Is there a difference in pause duration across age groups?* Figures 3 and 4 present the average pause duration for each cohort in histograms. As for unfilled pauses, the young cohort had an average unfilled pause duration of 952 milliseconds (*SD* = 217) and a median of 936 milliseconds, whereas the middle-aged cohort had an average duration of 1227 milliseconds (*SD* = 397) and a median of 1288 milliseconds, and the older cohort had an average duration of 1564 milliseconds (*SD* = 506) and a median of 1317 milliseconds. However, there is no main effect of group for the duration of unfilled pauses as shown by a Kruskal-Wallis test ($\chi^2(2) = 5.59$, $p = 0.061$). As for filled pauses, the young cohort had an average filled pause duration of 437 milliseconds (*SD* = 38) and a median of 427 milliseconds, whereas the middle-aged cohort had an average duration of 522 milliseconds (*SD* = 136) and a median of 557 milliseconds, and the older cohort had an average duration of 538 milliseconds (*SD* = 277) and a median of 422 milliseconds. However, there is no main effect of group for the duration of filled pauses as shown by a Kruskal-Wallis test ($\chi^2(2) = 1.57$, $p = 0.457$).
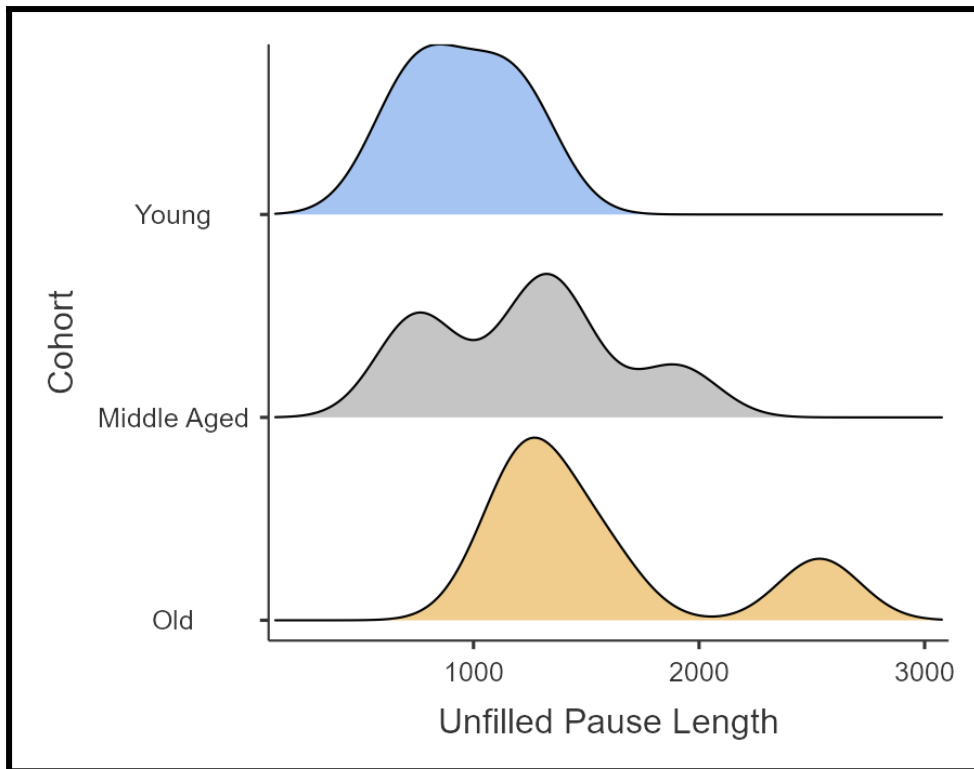
Figure 3: Histogram of the duration of unfilled pauses in milliseconds across the cohorts.
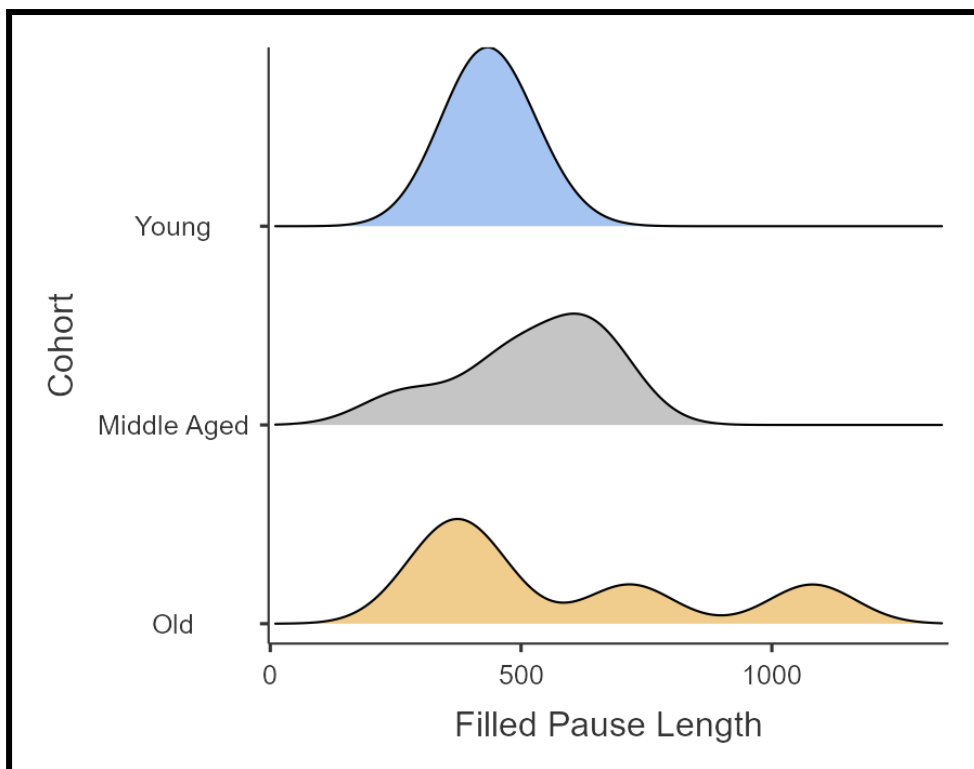

Figure 4: Histogram of the duration of filled pauses in milliseconds across the cohorts.

5.1.2. Speech errors

The fourth research question is *Is there a difference in the overall rate of speech errors across age groups?* In the current study, only 18 speech errors were found, and therefore no meaningful analysis of the rates of disfluencies can be performed. In sum, results regarding speech errors are inconclusive. However, descriptively, the rate of speech errors remained relatively constant, with the young cohort having an overall speech error rate of 0.24% (SD = 0.31%), the middle-aged a rate of 0.18% (SD = 0.16%), and the oldest cohort a rate of 0.34% (SD = 0.41%). The average rate across all cohorts is 0.25% (SD = 0.31%).

While the sample is exceedingly small and therefore inconclusive, some possible trends are suggested by the data. For younger adults, speech errors are most often seemingly caused by conceptual factors, that is, planning ahead and accidentally saying the incorrect word. For instance, when describing stimulus C3, participant Y2 accidentally produces the letter inside the orange node before mentioning the orange node itself. An overt repair reveals the error, as in (4).

(4) *blå (320.) vänster D<' (300.) eh(341,) orange D så E*

blue (320.) left D<' (300.) eh(341,) orange D so E

Similarly, in the aforementioned error produced by Y1 when describing stimulus C4 as shown in (3), the participant confused the starting point with the highest point, possibly since she is speaking about a node being *above* another node. Moreover, compared to the other cohorts, the younger cohort produced fewer hesitations following speech errors, more often repairing the error immediately after the incorrect word or an edit signal, as seen in (5) by participant Y4 describing stimulus C4.

(5) *sen till hö'< vänster om (482.) orange*

then to the ri'< left of (482.) orange

As for the middle-aged cohort, most notably, their repairs more often occurred in the middle of a word than after the word has been finished, compared to the other cohorts. For example, when describing stimulus S4, participant M1 produced the speech error in (6). M1 substituted *pink* for *red*, but repaired the utterance before completing the incorrect word.

(6) *sen går du (740.) upp (1304.) till (621.) rö'< ros:a*

then you go (740.) up (1304.) to (621.) re'< pin:k

Moreover, the middle-aged cohort produced speech errors both without hesitations, such as (6), and with hesitations, such as M4's description of C4 in (7), at equal rates. .

(7) *en kort sträcka åt väns'< (253.) åt höger*

a short way to the lef'< (253.) to the right

Additionally, middle-aged adults' substitutions were seemingly caused by associative and conceptual intrusions at equal rates. An example of an associative substitution is the error produced by M4 in (7). Left and right are antonyms and therefore associated, and this association likely causes the speech error.

In contrast, the older cohort was most likely to produce associative substitutions. For example, when describing stimulus S2, participant O2 produced the error in (8).

(8) *där var tre orange'< eh(181,) tre eh(269,) rosa*

there were three orange'< eh(181,) tre eh(269,) pink

Importantly, no orange node exists in the network, meaning the participant was unlikely to conceptualise one. Consequently, the error likely arised due to associations between words for colours. Moreover, the older cohort also produced more substitutions followed by hesitations than the other cohorts. For example, in (9), participant O5 described stimulus S4 and paused before commencing their repair.

(9) *svart till väns'< (337.) höger menar jag*

black to the lef'< (337.) right I mean

Lastly, each cohort produced similar amounts of repairs and edit signals in relation to their speech errors.

5.1.3. Speech Rate

The fifth research question is formulated as *Is there a difference in speech rate across age groups?* Figures 5 and 6 illustrate the data. As for articulation rate, the young cohort had an overall articulation rate of 219 syllables per minute ($SD = 33$) and a median of 226, the middle-aged cohort had a rate of 177 syllables per minute ($SD = 37$) and a median of 178, whereas the older cohort had an average rate of 179 syllables per minute ($SD = 40$) and a median of 168. There is no main effect of group as evidenced by a Kruskal-Wallis test ($\chi^2(2) = 3.85$, $p = 0.146$). As for the general rate of production, the young cohort had an average rate of 152 syllables per minute ($SD = 35$) and a median of 160, the middle-aged cohort had an average rate of 116 syllables per minute ($SD = 40$) and a median of 111, and the older cohort had an average rate of 112 syllables per minute ($SD = 36$) and a median of 107. There is no main effect of group as evidenced by a Kruskal-Wallis test ($\chi^2(2) = 2.99$, $p = 0.224$).
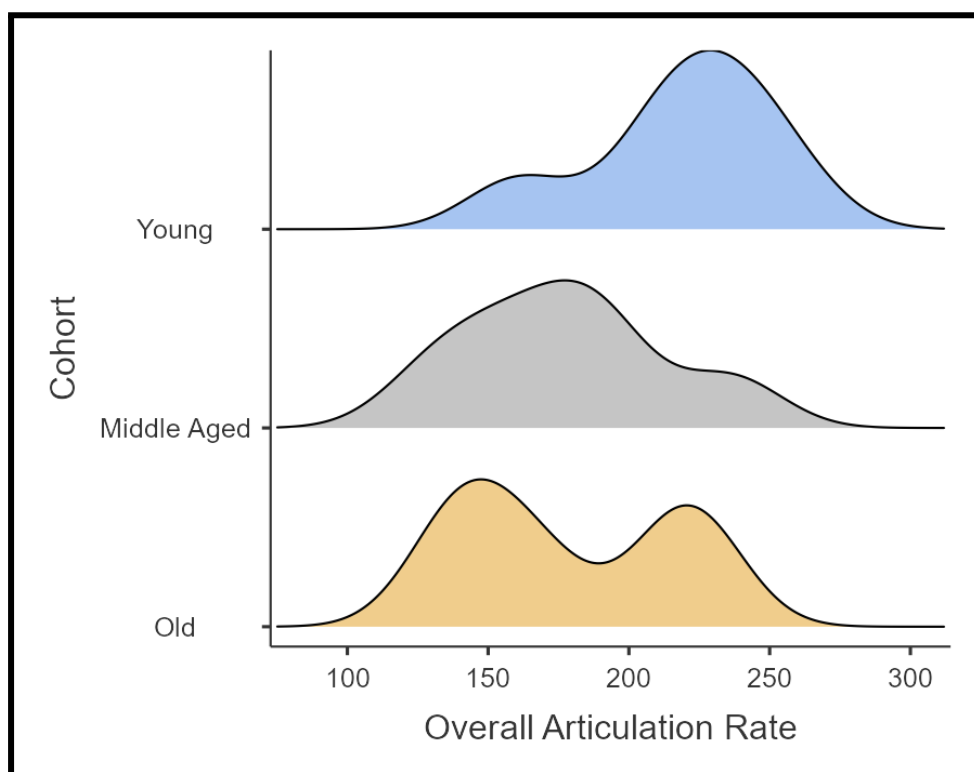
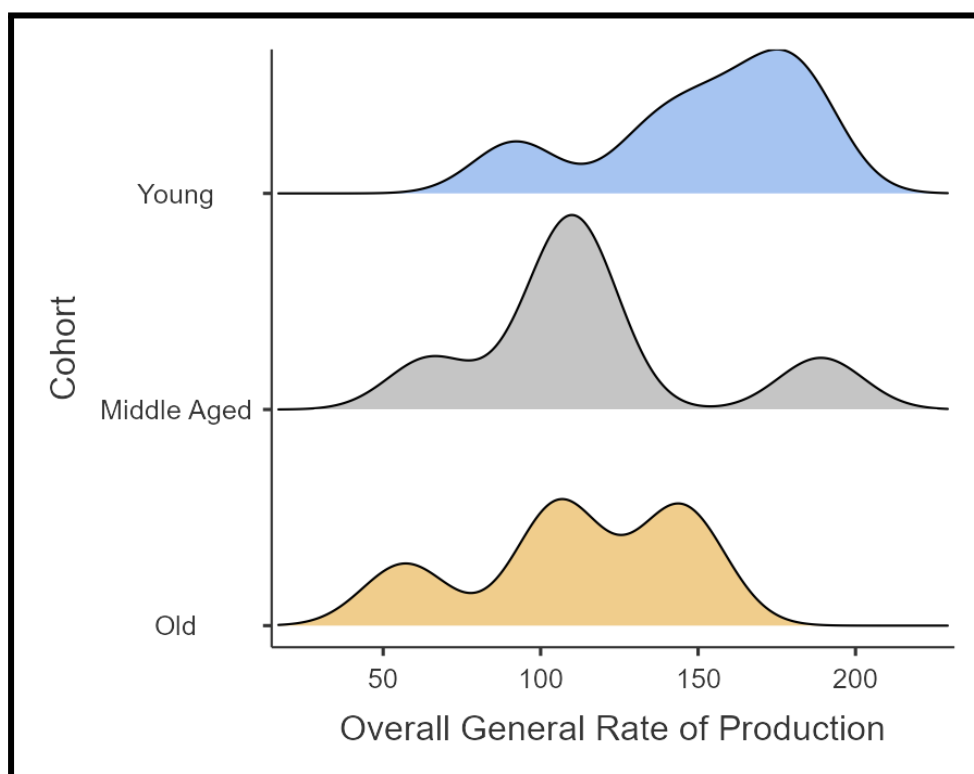*Figure 5: Histogram of the overall articulation rate in syllables per minute across the cohorts.*

Figure 6: Histogram of the general rate of production in syllables per minute across the cohorts.

5.2. Task Complexity and Fluency

The second half of the research questions is centred around the correlation between task complexity, age and fluency. Unfortunately, the method did not elicit sufficient data to properly analyse the correlation with age. Older adults either omitted the secondary tasks entirely or only performed certain parts of the tasks, such as O5 who omitted the green node task and only performed half of the orange node task, omitting to state the following letter in the alphabet. No older adult attempted every segment of each secondary task. Consequently, the effect of task complexity cannot be analysed for the oldest cohort, and this section only treats the young and middle-aged cohorts.

5.2.1. Disfluencies

The sixth research question is *Does the overall rate of disfluencies differ with task complexity across age groups?* On average, the young cohort had an overall disfluency rate of 16.9% (*SD* = 6.6%) for the simple material and a rate of 18.3% (*SD* = 6.7%) for the complex material. The middle-aged group had an average rate of 21.6% (*SD* = 5.9%) for the simple material and a rate of 22.54% (*SD* = 7.2%) for the complex material. To determine whether task complexity affected disfluency rates differently in the cohorts, a Wilcoxon test was performed. For the young cohort, the disfluency rate was not significantly different across the task complexity conditions ($Z = 7$, $p = 0.563$). Similar results were found for the middle-aged cohort ($Z = 8$, $p = 0.688$).

The seventh research question is *Do the rates of different types of disfluencies differ with task complexity and age?* Table 5 summarises the results. The different types did not occur in sufficient amounts to conduct an inferential analysis. Descriptively, most types were more frequent during the complex task, yet unfilled pauses decreased for the middle-aged cohort and interruptions decreased for the young cohort. Most notable were the increases in the rates of *eh* and repetitions. The average rate of *eh* increased from 1.42% (*SD* = 1.87%) in the simple task to 2.07% (*SD* = 1.86%) in the complex task for the young cohort, and from a rate of 1.15% (*SD* = 0.99%) in the simple task to 2.51% (*SD* = 1.58%) in the complex task for the middle-aged cohort. The average rate of repetitions increased from 0% *(SD* = 0%) in the

simple task to 0.04% (*SD* = 0.09%) in the complex material for the young cohort, and from a rate of 0.04% (*SD* = 0.1%) in the simple task to 0.38% (*SD* = 0.43%) in the complex task for the middle-aged cohort.

Table 5: Rate of each disfluency type for the simple and complex task respectively.

	Unfilled Pause	Eh	Uhm	Interruption	Prolongation	Repetition
Young Simple	12.99%	1.42%	0%	1.31%	0.85%	0%
Young Complex	13.08%	2.07%	0.21%	0.96%	1.37%	0.04%
Middle-aged Simple	17.24%	1.15%	0.95%	0.79%	1.43%	0.04%
Middle-aged Complex	14.43%	2.51%	1.64%	1.73%	1.85%	0.38%

The eighth research question is *Does pause duration differ with task complexity across age groups?* On average, the young cohort had an average unfilled pause duration of 814 milliseconds per pause (*SD* = 183) for the simple material and an average duration of 1034 milliseconds per pause (*SD* = 257) for the complex material. However, a Wilcoxon test of the duration of unfilled pauses for the young cohort in the simple task compared to the complex task showed no difference (*Z* = 1, *p* = 0.063). The middle-aged cohort had an average unfilled pause duration of 1358 milliseconds (*SD* = 687) for the simple material and an average duration of 1160 milliseconds (*SD* = 261) for the complex material. Again, there was no difference in the duration of unfilled pauses across the task complexity conditions (*Z* = 12, *p* = 0.313).

As for filled pauses, not all participants produced filled pauses during both levels of task complexity, and therefore some participants are excluded from the results. On average, the young cohort had an average filled pause duration of 386 milliseconds (*SD* = 85) for the simple task and an average duration of 454 milliseconds (*SD* = 49) for the complex task. However, a Wilcoxon test of the duration of unfilled pauses for the young cohort in the simple task compared to the complex task showed no significant difference (*Z* = 2, *p* = 0.750). The middle-aged cohort had an average filled pause duration of 623 milliseconds (*SD* = 293) for the simple material and a duration of 523 milliseconds (*SD* = 163) for the complex material. Again, there was no difference (*Z* = 9, *p* = 0.813).

5.2.2. Speech Rate

The ninth research question is *Does speech rate differ with task complexity across age groups?* On average, the young cohort had an articulation rate of 231 syllables per minute (*SD* = 39) for the simple task and 212 syllables per minute (*SD* = 27) for the complex task. However, a Wilcoxon test showed that there was no difference (*Z* = 17*, p* = 0.219). On average, the middle-aged cohort had an articulation rate of 179 syllables per minute (*SD* = 40) for the simple task and 176 syllables per minute (*SD* = 29) for the complex task. For the middle-aged cohort, there was no significant difference in articulation rate across task complexity conditions (*Z* = 13*, p* = 0.688).

As for the general rate of production, on average, the young cohort had an average rate of 165 syllables per minute (*SD* = 46) for the simple task and a rate of 145 syllables per minute (*SD* = 30) for the complex task. However, there was no significant difference (*Z* = 17*, p* = 0.219). On average, the middle-aged cohort had a general rate of production of 124 syllables per minute (*SD* = 41) for the simple task and a rate of 119 syllables per minute (*SD* = 33) for the complex task. Again, there was no significant difference (*Z* = 12.5*, p* = 0.752).

6. Discussion

6.1. Age and Fluency

The current study finds no statistically significant relationship between age and the overall rate of disfluency, different types of disfluencies, the rate of speech errors, articulation rate or general rate of production. Consequently, fluency does not appear to vary significantly with age, and the speech production process seemingly remains intact throughout the adult lifespan. However, certain trends can be observed. Importantly, however, these are not conclusive results.

While the overall rate of disfluencies seems to vary on a mostly individual basis, as evidenced by figure 2, certain types of disfluencies are more frequent in certain cohorts. In particular, repetitions appear to be more frequently produced by the older cohort. In total, the older cohort produced 19 repetitions across 1913 syllables, a rate of 1.06%, and a range of 2-8 repetitions per speaker. In contrast, the younger cohort produced only 1 repetition across 3202 syllables, a rate of 0.02%. Interestingly, the middle-aged cohort produced 10 repetitions across 3207 syllables, a rate of 0.21%, and a range of 0-6 repetitions per speaker. Although

the sample is too small to draw significant conclusions, it suggests that repetitions might increase with age. However, for prolongations, the opposite is true. The older cohort produced 9 prolongations in total with a range of 0-3 for the individual speakers, and an overall rate of 0.68%. In contrast, the youngest cohort produced a total of 44 with a range of 1-13 for the individual speakers, and a rate of 1.3%. However, the middle-aged cohort produced 54 prolongations with a range of 3-22 for the individual speakers, and an overall rate of 1.64%. While the results are inconclusive, the findings suggest that older adults produce fewer prolongations yet more repetitions than younger adults. If this is the case, it means that hesitation strategies vary across the cohorts. There are two main possible explanations. The first one is sociolectal. Since the study is cross-sectional, it is entirely possible that the previous or current linguistic environment of a cohort is different from the others, meaning that one cohort might simply subconsciously prefer one hesitation type over another. The second is cognitive. Interestingly, prolongations are, as mentioned, considered the least disfluent disfluency type while repetitions are at times considered to be a stuttering-like disfluency (Samani et al., 2017; Weathersby, 2016). Consequently, normal speech might become naturally more stuttering-like in old age. If this is the case, diagnostic definitions of stuttering speech must be careful to take natural changes in rates of repetitions into account when investigating older speakers.

Moreover, even though the duration of unfilled pauses does not differ significantly across age groups, there is a trend for older speakers to produce longer pauses. The average pause duration per cohort increases by circa 300 milliseconds from the young to the middle-aged to the old cohort respectively, in a near linear fashion. Figure 7 demonstrates this relationship.
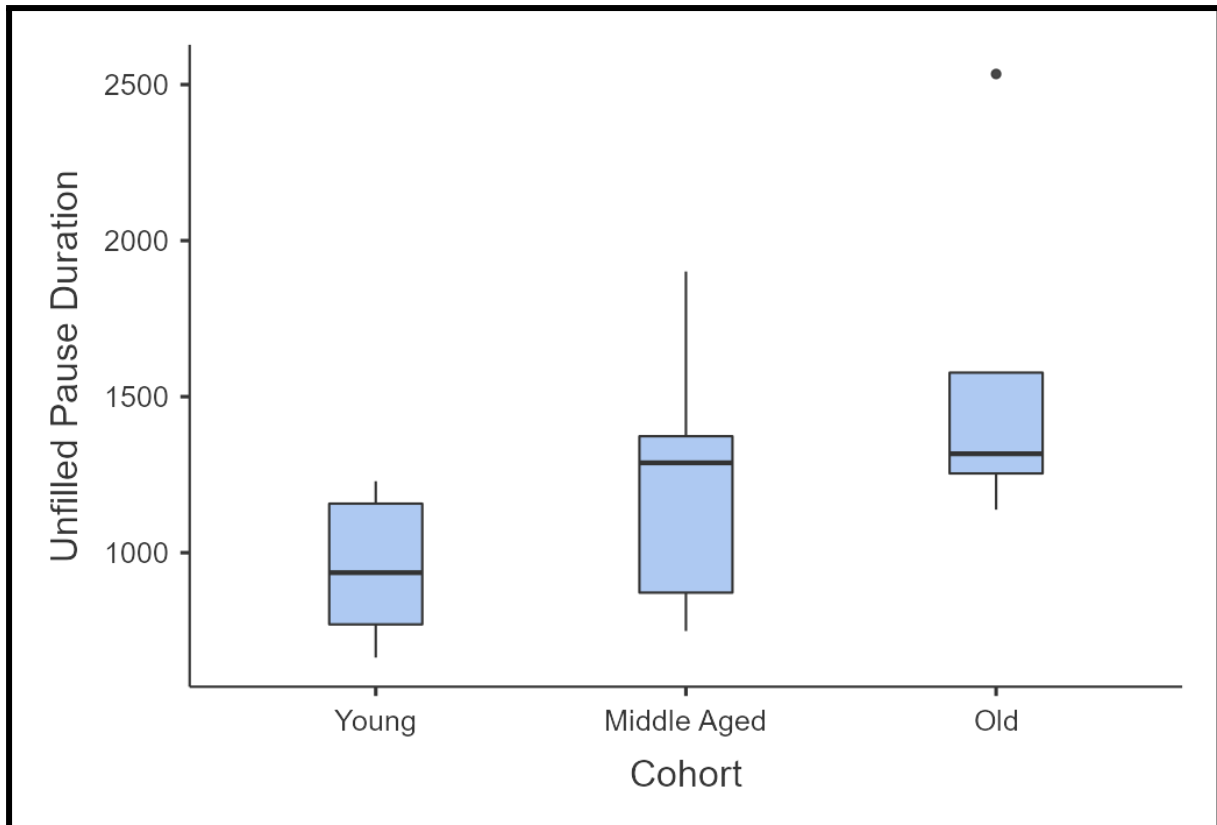
Figure 7: A boxplot of the duration of unfilled pauses in milliseconds per cohort.

This trend might arise due to an increased need for conceptualisation for older speakers. Since the measure utilised for this study includes extra-clausal pauses, most unfilled pauses likely arise due to the need to conceptualise. The task performed by the participants is highly reliant on memory, and consequently an abnormal amount of conceptualisation is needed compared to naturalistic speech, since the speaker must retrieve extensive information about the stimulus in order to describe it. Since older adults have impaired visuospatial memory compared to younger adults (Sandson et al., 1987), older participants might require, on average, longer time to retrieve information relevant to the task. However, another explanation is sociolectal. Since the study does not exclude extra-clausal pauses, delimitative pauses are included in the measure. Subjectively, the speech of young individuals can be impressionistically perceived as more fast paced and with more changes of topic, meaning there is greater need for holding the floor techniques such as avoiding long extra-clausal pauses. Consequently, younger adults might have embodied shorter pauses as a sociolect even during monologues, whereas older speakers might be used to a slower pace which allows for longer extra-clausal pauses. However, this is simply a hypothetical explanation.

Furthermore, while the sample of speech errors is exceedingly small, some trends were observed. Most notably, older adults produced more errors due to presumably associative intrusions than younger adults. Arbuckle et al. (2000) argue that older adults have less inhibitory control than younger adults, resulting in them conveying more irrelevant information during discourse. Hypothetically, an increase in associative speech errors might be caused by a physiological incapability to inhibit associative intrusions. Contrastingly, the younger adults produce more conceptual intrusions, suggesting they are more prone to swapping the location of two words in an utterance. Thus, younger speakers might be more susceptible to intrusions during grammatical encoding than older participants. Moreover, older adults also produced more speech errors followed by hesitations before the repair began than younger speakers. This does not suggest a delay in speech monitoring since the error is already noticed before the pause, yet it does suggest more time is needed to plan at some point in the speech production system since it takes longer time to amend the utterance. However, it should again be stressed that the findings are based on a small sample and consequently inconclusive, and future research is therefore direly needed to test the results.

Moreover, while neither speech rate measure differed significantly across the age groups, the younger cohort spoke notably faster than the older age groups. As for the general rate of production, the speech rate is likely faster since younger adults produce shorter unfilled pauses than the older cohorts. As for the articulation rate, older adults have less mobility and strength in the tongue (Magalhães Junior et al., 2014), meaning that articulation might be faster in younger adults for physiological reasons. However, the difference could yet again be sociolectal, possibly if younger adults are used to shorter speaking turns due to a faster paced discourse, meaning there is an increased need to produce the utterance quickly. Yet again, this explanation is purely hypothetical.

Additionally, while no significant relationship between any of the dependent variables and task complexity was found, younger adults produced more overall disfluencies and longer pauses for the complex task than the simple compared to the middle-aged cohort. As discussed, pause duration is likely caused by an increased need to reflect on the stimulus. Consequently, this suggests that younger adults found the complex task harder in relation to the simple task than the middle-aged cohort. Possibly, middle-aged participants have greater experience, both as speakers and in general, meaning they require less time to conceptualise their answers for the complex task. This also explains why there is greater individual variance

for the middle-aged speakers compared to the younger speakers, as evidenced by the standard deviations. In the first 20 years of life, most individuals in the study will have relatively similar experiences since they all attended some form of Swedish education. In contrast, in adulthood, individual experiences will vary more greatly depending on career choices, education, relationships, etc. Consequently, if the broad term *experience* is the reason for the greater individual variation in relation to the effects of task complexity on fluency, the middle-aged speakers should produce more varied results than the younger speakers, as is also the case.

6.2. Findings In Relation to Previous Research

First of all, the study finds no notable increase nor decrease in overall disfluency rates in semi-spontaneous speech, supporting the findings of Duchin & Mysak (1987), Keszler & Bóna (2019), Leeper & Culatta (1995), Searl et al. (2002) and Weathersby (2016) yet contrasting with the findings of Bortfeld et al. (2001), Horton et al. (2010) and Moscoso del Prado Martìn (2016). Interestingly, Keszler & Bóna (2019) note that while no significant correlation between age and disfluency rates can be found in their data, the most notable increase is in pause duration and frequency and propose that this might be due to physiological pauses for breathing. The current study removed physiological pauses and found no notable increase in the rate of unfilled pauses yet a notable, although insignificant, increase in unfilled pause duration was present. This corroborates Keszler & Bóna's (2019) findings. However, as discussed, these pauses might be sociolectal, or even due to the nature of the task relying on visuospatial memory, meaning no definitive conclusions about the speech production process can be drawn from them.

Second, the study finds no statistically significant relationship between age and the rates of different disfluency types, yet older adults produced more repetitions while younger adults produced more prolongations. Consequently, the study does not support the findings of Horton et al. (2010), who argue that the rate of filled pauses increases overall with age, while the rate of *eh* increases and the rate of *uhm* decreases. One possible explanation of the discrepant results supports the *filler-as-word* view, since true words are language specific and therefore undoubtedly vary in frequency and use between English, the language presumably studied by Horton et al. (2010), and Swedish. If this is the case, it indicates that taking the

filler-as-symptom view was a mistake in the current study that possibly skewed the results by not counting filled pauses as syllables.

Third, while speech rate was not statistically significantly different across age groups in the current study, younger speakers still spoke notably faster than the middle-aged and older cohorts on both measures. Since all previous studies mentioned in this thesis find a significant decrease in speech rate with age, the insignificant results in the current study are likely caused by a small sample size. However, previous studies find either that young and middle-aged speakers have similar speech rates whereas only the oldest speakers have reduced rates (Samani et al., 2017) or that speech rate decreases linearly with age (Duchin & Mysak, 1987; Leeper & Culatta, 1995). In contrast, the current study finds that younger participants speak at a faster rate, whereas middle-aged and old speakers have similar rates. While the findings are insignificant, they suggest that a decline in speech rate might be relatively language specific, and perhaps more influenced by sociolectal factors than previously thought.

6.3. Future Studies

Due to the statistically insignificant results found in the current study and the conflicting results found in previous studies, further research is needed to determine if fluency and age are correlated. The current study highlights both problems and possibilities for future studies.

While previous studies of aging and speech errors have refrained from analysing spontaneous speech, the current study proves that such an analysis is possible. With the stimuli from Melinger & Kita (2007), substitutions are reliably coded circa once per 300 words, compared to the average expected frequency of once per 1,000 words for all speech errors, as discussed in the theory. However, the current study does not elicit a sufficient quantity of data due to the small number of participants, meaning that a meaningful analysis is not possible. Although, with a larger sample size, an analysis of speech errors in semi-spontaneous speech is feasible with this method.

However, the method is not ideal for studies of disfluencies. As mentioned, several participants did not produce traditional clauses, instead opting to produce telegraph-like speech. Consequently, excluding extra-clausal disfluencies was not possible, meaning the

disfluency rates are skewed. Moreover, pause duration is likely influenced by age-related deficits in visuospatial memory, and separating possible limitations in memory from possible limitations in the speech production process is impossible with the current method. Therefore, stimuli that are not reliant on memory are likely to be preferable to studies investigating the correlation between age and fluency.

Additionally, to investigate the effects of task complexity, simpler stimuli would be advantageous. During debriefing, most participants stated that the complex stimuli were harder to describe than the simple ones, yet several stated that there was no difference, and a few even claimed that the simple stimuli were harder. Consequently, it is possible that the difference between the two levels was not great enough to properly investigate task complexity. Similarly, older participants appeared to find the complex level too difficult since they tended to omit secondary tasks. Future studies should therefore consider making both levels of complexity simpler, and to increase the difference in difficulty between the simple and the complex task to elicit greater differences.

7. Conclusions

The current study finds no statistically significant relationship between age and the overall rates of disfluency, the different types of disfluencies, pause duration, speech errors, nor speech rate. While the results are inconclusive, there is a trend for older adults to produce more repetitions and fewer prolongations, to produce longer unfilled pauses, to produce more associative than conceptual substitutions, and to speak more slowly than the younger adults. The middle-aged cohort is in between the youngest and the oldest cohort regarding repetitions, prolongations, pause duration and substitutions, yet have similar speech rates to the oldest cohort. No definitive explanation can be provided for these inconclusive trends, yet they could be caused by cognitive or sociolectal differences. As for task complexity, no significant difference is found between task complexity and the overall rate of disfluency, the different types of disfluencies, pause duration nor speech rate. However, younger adults produce notably longer unfilled pauses than the middle-aged cohort for the complex task compared to the simple task, suggesting they find it more straining. Possibly, the increased linguistic and general life experience of older speakers render them less affected by task complexity.

To conclude, until the issue can be examined in a larger sample, it appears that regardless of task complexity, fluency remains intact throughout the adult lifespan.

References

Amino, K., & Osanai, T. (2015). Cross-language differences of articulation rate and its transfer into Japanese as a second language. *Forensic Science International*, *249*, 116-122. DOI: 10.1016/j.forsciint.2015.01.029

Arbuckle, T.Y., Nohara-Leclair, M., & Pushkar, D. (2000). Effect of off-target verbosity on communication efficiency in a referential communication task. *Psychology and Aging*, *15*(1), 65-66. DOI: 10.1037/0882-7974.15.1.65

Batliner, A., Kießling, A., Burger, S., & Nöth., E. (1995). Filled pauses in spontaneous speech. *Proceedings of 13th International Congress of Phonetic Sciences*, *3*, 472-475. DOI: 10.22028/D291-25176

Betz, S., & Wagner, P. (2016). Disfluent lengthening in spontaneous speech. In Jokisch, O. (Ed.), *Elektronische Sprachsignalverarbeitung*. (pp. 135-144). TUDPress.

Bortfeld, H., Leon, S.D., Bloom, J.E., Schober, M.F., & Brennan, S.E. (2001). Disfluency rate in conversation: Effects of age, relationship, topic, role, and gender. *Journal of Language and Speech*, *44*(2), 123-147. DOI: 10.1177/00238309010440020101

Clark, H.H., & Fox Tree, J.E., (2002). Using uh and uhm in spontaneous speaking. *Cognition*, *84*, 71-111. DOI: 10.1016/s0010-0277(02)00017-3

Coupé, C., Mi Oh, Y., Dediu, D., & Pellegrino, F. (2019). Different languages, similar encoding efficiency: Comparable information rates across the human communicative niche. *Science Advances*, *5*(9), eaaaw2594. DOI: 10.1126/sciadv.aaw2594

Duchin, S.W., & Mysak, E.D. (1987). Disfluency and rate characteristics of young adult, middle-aged, and older males. *Journal of Communication Disorders*, *20*, 245-257. DOI: 10.1016/0021-9924(87)90022-0

Drayna, D., Kilshaw, J., & Kelly, J. (1999). The sex ratio in familial persistent stuttering. *The American Journal of Human Genetics*, *65*(5), 1473-1475. DOI: 10.1086/302625

Fraundorf, S.H., Arnold, J., & Langlois, V.J. (22-02-2018). *Disfluency*. Oxford Bibliographies. https://www.oxfordbibliographies.com/view/document/obo-9780199772810/obo-9780199772810-0189.xml

Gollan, T.H., & Goldrick, M. (2019). Aging deficits in naturalistic speech production and monitoring revealed through reading aloud. *Psychology and Aging*, *34*(1), 25-42. DOI: 10.1037/pag0000296.

Graziano, M. & Gullberg, M. (2018). When speech stops, gesture stops: evidence from developmental and crosslinguistic comparisons. *Frontiers in Psychology*, *9*, 879. DOI: 10.3389/fpsyg.2018.00879

Harley, T.A. (2006). Speech errors: psycholinguistic approach. In K. Brown. (Ed.). *Encyclopedia of Language & Linguistics* (pp. 739-745). Elsevier.

Hilton, N.H., Schüppert, A., & Gooskens, C. (2011). Syllable reduction and articulation rates in Danish, Norwegian and Swedish. *Nordic Journal of Linguistics, 34*(2), 215-237. DOI: 10.1017/S0332586511000175

Hooper, C. R., & Cralidis, A. (2009). Normal changes in the speech of older adults. You've still got what it takes, it just takes a little longer! *Perspectives on Gerontology*, *14*(2), 47-56. DOI: 10.1044/gero14.2.47

Horton, W.S., Spieler, D.H., & Shriberg, E. (2010). A corpus analysis of patterns of age-related change in conversational speech. *Psychology of Aging*, *25*(3), 708-713. DOI: 10.1037/a0019424

Jacewicz, E., Fox, R.A., & O'Neill, C. (2009). Articulation rate across dialect, age and gender. *Language Variation and Change*, *21*, 233-256. DOI: 10.1017/S0954394509990093

Johnson, W. (1961). Measurements of oral reading and speaking rate and disfluency of adult male and female stutterers and nonstutterers. *The Journal of Speech and Hearing Disorders*, *7*, 1-20.

Kavé, G., & Goral, M. (2017). Do age-related word retrieval difficulties appear (or disappear) in connected speech? *Aging Neuropsychology and Cognition*, *25*, 508-527. DOI: 10.1080/13825585.2016.1226249

Kendall, T. (2013). *Speech Rate, Pause, and Sociolinguistic Variation: Studies in Corpus Sociophonetics*. Palgrave Macmillan.

Keszler, B., & Bóna, J. (2019) Pausing and disfluencies in elderly speech: Longitudinal case studies. In R.L. Rose., & R. Eklund (Eds.), *Proceedings of DiSS 2019, The 9th Workshop on Disfluency in Spontaneous Speech*, (pp. 67-70). KTH Royal Institute of Technology.

Leeper, L.H., & Culatta, R. (1995). Speech fluency: effect of age, gender and context. *Folia Phoniatrica et Logopedia*, *47*, 1-14. DOI: 10.1159/000266337

Levelt, W.J.M. (1989). *Speaking: From Intention to Articulation*. MIT Press.

Lickley, R.J. (2015). Fluency and disfluency. In M.A. Redford (Ed.) *The Handbook of Speech Production*, (pp. 445–474). Wiley-Blackwell.

Lickley, R.J. (2018). *Disfluency in Typical and Stuttered Speech.* DOI: 10.17469/O2103AISV000019

Litosseliti, L. (Ed.). (2010). *Research Methods in Linguistics*. Continuum International Publishing Group.

Magalhães Junior, H.V., Tavares, J.C., Magalhães, A.A.B., Galvão, H.C., & Ferreira, M.A.F. (2014). Characterization of tongue pressure in the elderly. *Audiology - Communication Research*, *19*(4), 375-379. DOI: 10.1590/S2317-64312014000400001401

MacGregor, L. J., Corley, M., & Donaldson, D.I. (2010). Listening to the sound of silence: disfluent silent pauses in speech have consequences for listeners. *Neuropsychologia*, *48*(14), 3982-3992. DOI: 10.1016/j.neuropsychologia.2010.09.024

MacKay, D.G., & James, L.E. (2004). Sequencing, speech production, and selective effects of aging on phonological and morphological speech errors. *Psychology and Aging*, *19*(1), 93-107. DOI: 10.1037/0882-7974.19.1.93

Matthews, P. (2007). *Unfilled pause*. Oxford Reference. https://www.oxfordreference.com/view/10.1093/acref/9780199202720.001.0001/acref-97801 99202720-e-3534

Melinger, A., & Kita, S. (2007). Conceptualisation load triggers gesture production. *Language and Cognitive Processes*. *22*(4), 473-500. DOI: 10.1080/01690960600696916

Moscoso del Prado Martín, F. (2017). Vocabulary, grammar, sex, and aging. *Cognitive science*, *41*(4), 950–975. DOI: 10.1111/cogs.12367

Rose, R. (2017). Silent and filled pauses and speech planning in first and second language production. In R. Rose., & R. Eklund (Eds.). *TMH-QPSR* (pp. 49-52). Stockholm Royal Institute of Technology.

Sandson, J., Obler, L.K., & Albert, M.L. (1987). Language changes in healthy aging and dementia. In S. Rosenberg (Ed.). *Advances in Applied Psycholinguistics*, (pp. 264-291). Cambridge University Press.

Samani, M., Abnavi, F., & Ghasisin, L. (2017). Do older adults experience changes in their speech fluency? Some evidence from Iranian elderly people. *Journal of Clinical Gerontology & Geriatrics*, *8*(4), 127-133. DOI: 10.24816/jcgg.2017.v8i4.05

Statistiska centralbyrån, (2018-04-26). *Störst folkökning att vänta bland de äldsta*. Statistikmyndigheten. https://www.scb.se/hitta-statistik/statistik-efter-amne/befolkning/befolkningsframskrivningar/ befolkningsframskrivningar/pong/statistiknyhet/sveriges-framtida-befolking-20182070/

Searl, J.P., Gabel, R.M., & Fulks, J.S. (2002). Speech disfluency in centenarians. *Journal of Communication Disorders*, *35*, 383-392. DOI: 10.1016/S0021-9924(02)00084-9

Van Borsel, J., & De Maesschalck, D. (2008). Speech rates in males, females, and male-to-female transsexuals. *Clinical Linguistics & Phonetics*, *22*(9), 679-685. DOI: 10.1080/02699200801976695

Verhoeven, J., De Pauw, G., & Kloots, H. (2004). Speech rate in a pluricentric language: a comparison between dutch in Belgium and the Netherlands. *Language and Speech*, *47*, 297-308. DOI: 10.1177/00238309040470030401

Volpi, E., Nazemi, R., & Fujita, S. (2004). Muscle tissue changes with aging. *Current Opinion in Clinical Nutrition and Metabolic Care*, *7*(4), 405-410. DOI: 10.1097/01.mco.0000134362.76653.b2

Vousden, J.I., & Maylor, E.A. (2006). Speech errors across the lifespan. *Language and Cognitive Processes*, *21*, 48-77. DOI: 10.1080/01690960400001838

Warren, P. (2012). *Introducing Psycholinguistics*. Cambridge University Press.

Weathersby, A. (2016). *Disfluency and Ageing: A Study of Healthy Older Speakers of New Zealand English*. (MSc thesis), University of Canterbury, Christchurch.

Appendix A

The script used during each elicitation, including questions regarding the participant's background. Note that since the script is meant to be read aloud it is written in a relatively vernacular style.

Tack för att du vill delta i min studie. Jag kommer följa ett manus för att se till så att alla deltagare får exakt samma instruktioner. Först och främst får du ett samtyckesformulär. Läs igenom detta noggrant och ställ gärna frågor. Om du känner dig bekväm med villkoren får du skriva under pappren. Det är två identiska kopior, en till mig och en till dig.
Här är formuläret.
...
Tack. Som du läste på formuläret är den första delen av studien en kortfattad intervju som vi tar muntligt, och jag kommer skriva ner dina svar. Låter det bra?
...
Okej. Vad är ditt juridiska kön?
...
Hur gammal är du?
...
Vad är din högsta avklarade examen? Exempelvis högstadie, gymnasie, kandidatnivå, et cetera.
...
Vad är ditt modersmål?
...
Vad skulle du själv säga att du har för dialekt?
...
I vilket landskap tillbringade du mest tid i under din uppväxt?
...
Det var alla frågor jag hade, och vi börjar nu med den första delen av själva experimentet.
...
Så, vi kommer att börja med en träningsrunda som vi inte spelar in för att se till så att allt går bra. Du kommer att få se en bild i femton sekunder som föreställer ett vägnätverk fyllt med färgglada cirklar som ser ut nästan som en karta. Efter femton sekunder tas bilden bort. Du ska då berätta hur nätverket ser ut utifrån cirklarna, där du börjar med den cirkeln längst ner som en svart pil pekar på, berättar vilken färg den är, och vilken riktning du ska ta för att komma till nästa cirkel samt vilken färg den är, och så vidare tills du beskrivit hela nätverket. Du får inte hoppa över cirklar, utan du måste försöka ta en cirkel i taget. Som sagt börjar vi med en träningsrunda. Bilden visas på datorn och försvinner automatiskt efter 15 sekunder, och det är då du ska börja beskriva den. Har du förstått, eller har du kanske några frågor?
...
Nu kommer du att få göra en till träningsrunda som är lite annorlunda. Du kommer först att få titta på ett till vägnätverk i femton sekunder. Du ska beskriva detta nätverk på samma sätt

som du beskrev den tidigare, men i detta vägnätverk finns också två nya cirklar, nämligen en grön och en orange. När du kommer till den gröna cirkeln vill jag att du berättar var den cirkeln befinner sig i förhållande till cirkeln du började på, alltså fågelvägen. Sedan ska du fortsätta beskriva nätverket från den gröna cirkeln. När du kommer till den orangea cirkeln vill jag att du berättar vilken bokstav som är skriven inuti den, samt vilken bokstav som följer i alfabetet. Sedan fortsätter du beskriva nätverket. Vi börjar med en träningsrunda. Har du förstått, eller har du några frågor?

...

Nu kan vi påbörja det riktiga experimentet. Precis som under övningarna kommer du att få titta på en bild i femton sekunder. Sedan ska du beskriva hur vägnätverket ser ut utifrån cirklarna. Detta upprepar vi åtta gånger på raken, med åtta olika bilder. Fyra av bilderna kommer innehålla gröna och orangea cirklar, och fyra fem kommer inte att göra det. Det är viktigt att du försöker göra så få misstag som möjligt när du beskriver bilden. Efter att du fått se en bild kommer jag ta datorn så att jag kan kolla på bilden medan du beskriver den. Jag kommer inte att säga något om jag inte behöver, utan kommer bara sitta tyst. När du är färdig med beskrivningen får du direkt en ny bild, tills vi är färdiga. Jag börjar nu spela in ljud, om du är redo?

...

...

...

Nu är vi färdiga med experimentet och jag stänger av inspelningen. Om du vill kan jag nu förklara mer djupgående vad det är för experiment du har gjort och vad syftet är. Är du intresserad av att höra mer om vad studien går ut på?
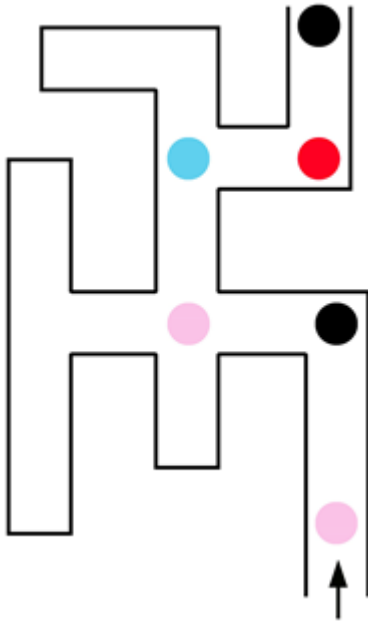
Appendix B

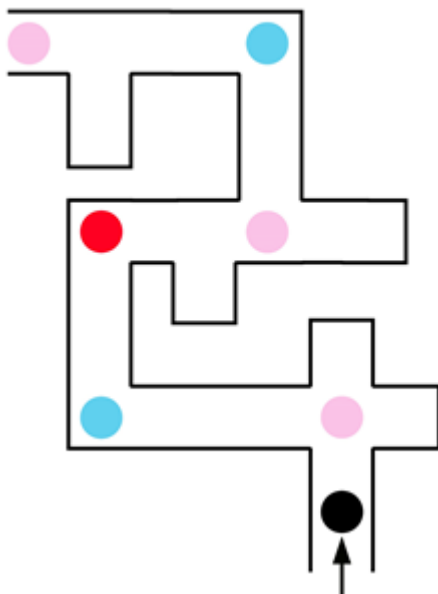The stimuli created for the study.

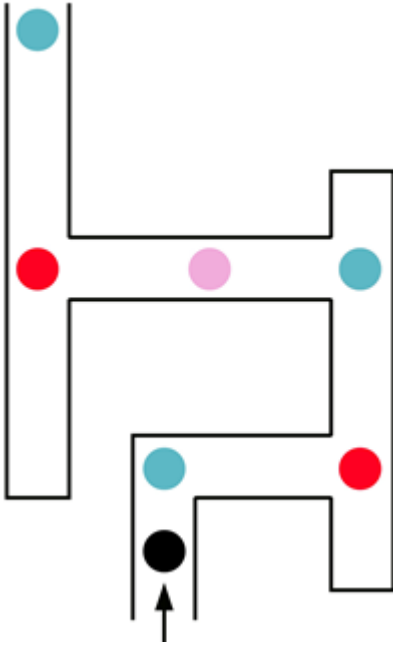*Figure 8: Stimulus S1.*

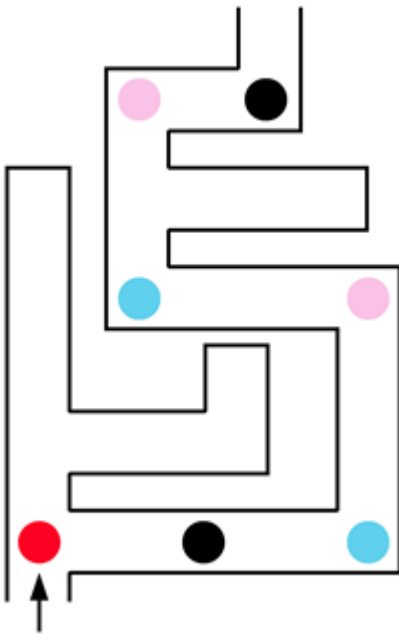*Figure 9: Stimulus S2.*

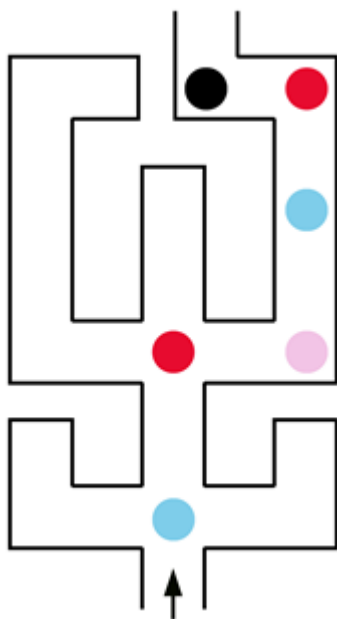*Figure 10: Stimulus S3.*

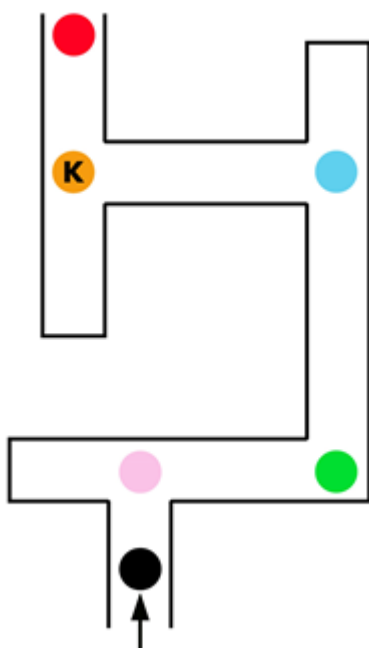*Figure 11: Stimulus S4.*

Figure 12: Stimulus S5.
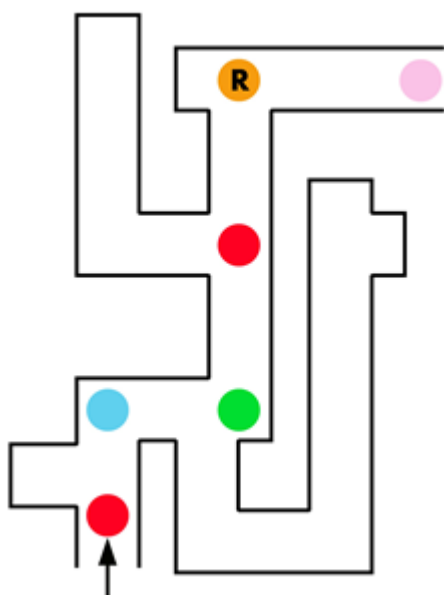

Figure 13: Stimulus C1.
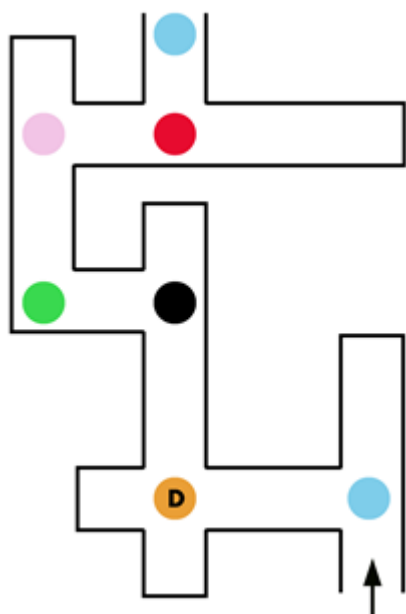
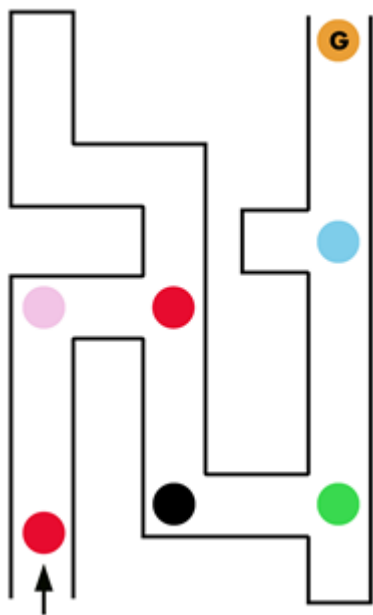*Figure 14: Stimulus C2.*

Figure 14: Stimulus C3.
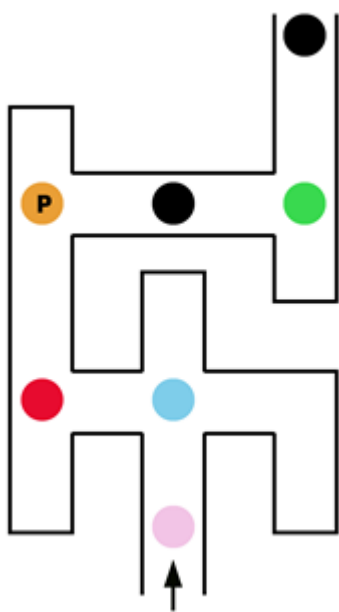
Figure 15: Stimulus C4.

Figure 16: Stimulus C5.

Appendix C

The consent form used for the study.

SAMTYCKE TILL ATT DELTA I FORSKNINGSPROJEKTET

Kandidatstudent Harry Polfeldt, Språk- och litteraturcentrum, Lunds universitet

Du deltar i en studie som är en del av examinationen i kursen ALSK13 som är en kurs på kandidatnivå inom språkvetenskap på Språk- och litteraturcentrum på Lunds universitet. Studien genomförs av kandidatstudent Harry Polfeldt med professor Marianne Gullberg som handledare. Syftet är att undersöka hur talat språk skiljer sig mellan olika åldersgrupper.

Studien består av att du som deltagare till att börja med får besvara frågor om ditt juridiska kön, ditt födelseår, din utbildning och din dialekt. Efter frågorna får du se tio olika bilder i två kategorier som du får beskriva från minnet. Dessa beskrivningar sker under ljudinspelning.

All data är anonym i den färdiga uppsatsen. Endast kandidatstudent Harry Polfeldt och handledare Marianne Gullberg har tillgång till datan. All data förvaras på kandidatstudent Harry Polfeldts dator tills studien är avslutad. Därefter kommer den att raderas.

Ditt deltagande är helt och hållet frivilligt och du kan när som helst avbryta deltagandet och återkalla ditt samtycke, även efter experimentet är genomfört.

Med din underskrift nedan intygar du att du har läst och förstått informationen ovan samt att du frivilligt samtycker till deltagande. Du kan dra tillbaka ditt samtycke eller avbryta deltagandet när som helst utan påföljder. Du behåller en kopia av detta samtyckesformulär.

Datum & Underskrift

_____

Kontaktpersoner

Harry Polfeldt
0701454144
harrypolfeldt@gmail.com

Marianne Gullberg
marianne.gullberg@ling.lu.se