Um or Um?

Language- and Speaker-specificity of Filled Pauses in Dutch and English

Jade van der Graaf

4169980

BA Thesis English Language and Culture, Utrecht University

Supervisors: Dr. Koen Sebregts & Dr. Willemijn Heeren

Second reader: Prof. Dr. Aoju Chen

February 2018

Abstract

Spontaneous speech is extremely disfluent, as speakers use all manner of hesitation when speaking. This study focuses on one type of hesitation: the filled pause, specifically *uh* and *um*. The English and Dutch speech of 20 native Dutch speakers is examined in order to establish the differences in use for these two languages. Previous research predicts transfer from a speaker's native language to an L2. The present study found transfer for three of the four aspects of filled pauses that were examined, namely: duration, number and type of filled pause. There was no transfer found for the first and second formants of the vowels used in the filled pauses. The fact that there is transfer to a certain extent is relevant for the domain of forensic linguistics, as this suggests that speakers remain recognisable to a degree when speaking a second language.

Index

1. Introduction

1.1 Hesitation in speech

Spontaneous speech is notoriously disfluent. A considerable amount of literature has been published on this disfluency. Maclay and Osgood (1959) define four types of hesitation in speech. The first is a repeat, which they describe as a repetition of any length that is semantically non-significant. For example, "I I saw a very very big boy" (Maclay & Osgood, 1959, p. 24). *I* is a repeat, but *very* is not, because in the case of *very*, the meaning of the sentence changes. A repeat can vary from a single phoneme to a string of words. The second hesitation type is a false start. This includes all incomplete or self-interrupted utterances. The third hesitation type defined by Maclay and Osgood is a filled pause, which is a pause in speech production that has been filled with sound. The final type of hesitation in speech is an unfilled pause, which is an abnormal hesitation in speech that could not be referred to one of the previous categories. The most common forms defined by Maclay and Osgood are silences of unusual length or non-phonemic lengthening of phonemes.

The present study will focus on filled pauses, as defined by Maclay and Osgood. An overview of the existing literature on this subject will lead to the matter of to what extent Dutch speakers use filled pauses differently when speaking their native or a foreign language, and, more specifically, to what extent they remain recognisable by these filled pauses when speaking a foreign language.

1.2 Filled pauses: uh **and** *um*

As previously mentioned, a filled pause is a pause in speech production that has been filled with sound. Two types of filled pauses are typically identified: *uh* and *um.* Wieling et al. (2014) describe these as consisting of a neutral vowel, often the schwa [ə], in an open syllable, which in the case of *um* is followed by a final labial nasal (Wieling et al., 2014, p.

200). Clark and Fox Tree (2002) describe many features of *uh* and *um*, and argue that they are not merely filled pauses, but actual English words, by which they mean linguistic units that have conventional phonological shapes and meanings, governed by the rules of syntax and prosody. They state that *uh* and *um* are planned for and formulated as part of utterances as any other word would be. Clark and Fox Tree describe three earlier views on *uh* and *um*. The first is that they are fillers as symptoms of problems detected by the speaker during speaking. The second view posits that they are fillers used as non-linguistic signals, for example for holding the floor. The third view is that *uh* and *um* are proper words, specifically, interjections (Clark & Fox Tree, 2002, p. 75). They also found that speakers use these words to announce delays in speaking, where *uh* indicates a short delay and *um* a long one. Clark and Fox Tree discovered, by looking at several large English corpora, that speakers plan for delays and decide where and how to pause their speech: *uh* and *um* are often cliticised onto preceding words, though never onto following words, and they state that the cliticisation of *uh* and *um* onto previous words is evidence that they are the initiation of an expected delay (Clark & Fox Tree, 2002, p. 101).

Researchers have compared the usage of *uh* and *um*, some finding a difference in use, and others finding no functional difference. These studies have been done on a variety of languages, which could account for some of the diversity in results. Clark and Fox Tree (2002) found that *uh* facilitated word recognition by listeners, but *um* did not (Clark & Fox Tree, 2002, as cited in Wieling et al, 2014, p. 201). Swerts (1998) found that *um* is used more often at the beginning of an utterance, and *uh* more often in medial positions (Swerts, 1998, p. 490). However, O'Connell and Kowal (2005) argue that there is no functional difference in the usage of *uh* and *um.* This is based on an analysis of six interviews with Hillary Clinton. Moreover, a review of previous research by Corley and Stewart (2008) showed that there is no evidence to suggest that speakers can control the production of *uh* or *um* (O'Connell &

Kowal, 2005; Corley & Stewart, 2008, as cited in Wieling et al., 2014, p. 202). These findings are called into question by a recent study by Fruehwald (2016), who found that speakers not only control their production of *uh* and *um*, but they consciously choose between the two types of filled pauses. Fruehwald states that the selection of either *uh* or *um* is a social variable. Various studies have found that women and younger speakers use *um* more than they use *uh*, and this tendency has been found for many Germanic languages, such as Dutch, German, and British and American varieties of English (Acton, 2011; Tottie, 2011; Laserna et al, 2014, Wieling et al, to appear, as cited in Fruehwald, 2016, p. 43).

Many theories exist on why speakers use filled pauses, and while there is, as yet, no unanimity, most researchers seem to agree on the following functions. The first is that speakers use *uh* and *um* to indicate to their interlocutor that they are searching for a word. Other functions of filled pauses that are generally agreed upon are deciding what to say next, wanting to keep the floor, and wanting to cede the floor (Clark & Fox Tree, 2002, p. 73). Lallgee and Cook (1969) suggest that if speakers feel pressure to keep speaking but have nothing specific to say, they will fall back on filled pauses (Lallgee & Cook, 1969, p. 24). However, Broen and Siegel (1972) concluded that speakers were able to minimise the use of filled pauses when put in a situation in which speaking correctly was thought to be important (Broen & Siegel, 1972, p. 229). Bortfeld, Leon, Bloom, Schober and Brennan (2001) carried out a study on whether age, gender, relationship or topic have any effect on disfluency rates in conversation. They found that older speakers produced more disfluencies than younger speakers. Men had a higher rate of disfluencies than women, while the same number of words were spoken. This was due mainly to a higher rate of fillers. Married pairs did not have more fluent conversations than pairs of strangers, and there were fewer disfluencies when talking about a familiar topic than an unfamiliar one.

While the present study is focused on speakers as opposed to listeners, the majority of

the literature concentrates on the effects of filled pauses on listeners. It is a common assumption that hesitations slow or impair a listener's understanding of speech. However, Fox Tree's (1995) study shows that speech disfluencies have different effects on comprehension depending on the type and location of the disfluency. For example, it takes longer to identify words following false starts, but not following repetitions. A false start at the beginning of a sentence leads to less processing trouble than a false start in the middle of a sentence. A study by Brennan and Schober (2001), in which participants were asked to select a geometric object described to them as quickly as possible, showed that hesitation markers were beneficial to comprehension, as listeners selected a target object faster after a filler was used in the stimulus sentence. This is attributed to listeners having more time to process the stimuli (Brennan & Schober, 2001, p. 282). Arnold, Fagnano and Tanenhaus (2003) carried out a study in which participants followed pairs of instructions to move objects on a screen, in order to determine whether disfluencies signal new information. They found that when encountering disfluent speech, listeners were more likely to expect a discourse-new referent (Arnold et al., 2003, p. 35). Similarly, Bosker, Quené, Sanders and de Jong (2014) demonstrated that listeners were more likely to expect a low-frequency word (as opposed to a high-frequency word) after a disfluency marker, attributing the disfluency to the speaker having trouble with lexical retrieval. A second experiment in the same study, this time with non-native speakers, showed that there was no anticipation of low-frequency words after a disfluency. Bosker et al. concluded that listeners change their predictive strategies to different speakers and levels of native-ness (Bosker et al. 2014, p. 104).

This overview of what filled pauses are and how they are used is necessary to understand the relevance of investigating a speaker's filled pauses. The present study delves into the filled pauses *uh* and *um* in order to investigate the degree to which speakers are recognisable when speaking a foreign language.

1.3 Language specificity and relevance to forensic phonetics

Most researchers agree that the realisation of hesitation markers is language-specific, though it is not clear what causes this. According to Künzel (1997), German speakers tend to use the vowel space between [ə], [ɛ] and [œ], while French speakers use [œ] or [ø] (Künzel, 1997, p. 51). Of the English hesitation markers [ɛ], [æ], [ə] and [m], speakers most frequently use [ə] (Maclay & Osgood, 1959, p. 24). In Dutch, filled pauses are realised as [əm] or [ə] (Swerts, 1998, p. 486). The vowel space of Dutch and English [ə] is described by Collins et al. (2006) and presented precisely in appendix A. According to van Donzel and Koopmans-van Beinum (1996), Dutch speakers commonly cliticise vocalic hesitation markers onto preceding words (van Donzel & Koopmans-van Beinum, 1996, p. 1029). A study by De Leeuw (2007), which compared filled pauses in English, Dutch and German, found that, of the three languages, Dutch speakers used filled pauses most often and Germans speakers the least. German and English speakers used more vocalic-nasal filled pauses, while Dutch speakers used more vocalic pauses (De Leeuw, 2007, p. 98-99).

The fact that the realisation of filled pauses is language-specific is of great importance to the domain of forensic phonetics. The most central aspect of forensic phonetics is speaker identification, or speaker recognition (Jessen, 2008, p. 671). According to Künzel (1997), individuals are quite consistent in using their respective personal variant of the hesitation sound with respect to the vocalic quality and the addition of the bilabial nasal. Künzel also mentions that the frequency of occurrence of filled pauses, the proportion of filled and unfilled pauses, and the lengthening of word-final segments is speaker-specific (Künzel, 1997, p. 51). This would mean that speakers would be able to be recognised to a certain extent by their realisation of hesitation markers. Baldwin and French (1990) mention that hesitation markers are likely to be transferred from a speaker's native language to their foreign language (Baldwin & French, 1990, as cited in De Leeuw, 2007, p. 93).

The present study builds on those by Künzel and Baldwin and French, and compares the speech of speakers in their native Dutch and L2 English, in order to measure whether the realisation of hesitation markers, and filled pauses in particular, is indeed transferred from native to foreign language. In the event of transfer, this would imply a degree of speaker recognition, a major tenet of forensic linguistics. A lack of transfer would suggest that speakers correctly use their target language's filled pauses. By comparing the filled pauses used by Dutch speakers in English and Dutch, the present study explores to what extent they are used differently, in regard to amount of filled pauses, their duration, the number of vocalic to vocalic-nasal pauses and their formant structure. The existing literature suggests that as speakers transfer their hesitation behaviour from their native language to their L2, their filled pauses in both languages will be similar, specifically pertaining to vowel quality.

2. Methodology

The filled pauses of 20 Dutch speakers, aged between 18 and 23, were compared in English and Dutch to determine potential differences in their realisation and usage in the two languages. This study made use of the LUCEA corpus (Orr et al., 2011), which is a speech corpus of students from University College Utrecht (UCU), an English-language international Liberal Arts and Sciences College in the Netherlands. The present study is similar to De Leeuw's (2007), as both compare filled pauses over languages. Thus, certain aspects of the method derive from her study. De Leeuw operationalised the realisation of filled pauses using four parameters: the number of filled pauses per minute of speech, the proportion of vocalic, vocalic-nasal and nasal pauses, the duration of the filled pauses, and their positioning. These parameters, save for the positioning of filled pauses, will be used in the present study. Positioning will be replaced by formant frequency, as this was deemed more informative with regards to speaker recognisability.

2.1 Participants

The LUCEA corpus was set up by Orr et al. as part of a longitudinal study on accent development. The students were recruited from the first-semester students in September 2010, 2011 and 2012 and were recorded 5 times over the course of their three-year undergraduate studies. For this study, however, only the first set of recordings from 2010 were used. The speakers are similar in age and education level. All were Dutch native speakers, with varying proficiency in English[1]. Ten female and ten male speakers were included.

2.2 Material

The recordings from the LUCEA corpus each consist of seven parts recorded in a single sound file. These seven parts include two read texts, sentences for prosodic analysis and intelligibility testing, prepared but not practiced speech on two topics (one formal and one informal), and free conversational speech. If English was not the speaker's L1, they were asked for a sample of speech in their native language as well (Orr et al., 2011, p. 1890). As the Dutch speech sample was prepared speech on an informal topic, the same segment was used in English, to make the comparison as reliable as possible. In most cases, the speakers spoke about their summer holidays or the introduction week of their study. They often spoke on the same topic in English and Dutch. The prepared speech was approximately 2.5 minutes in each language and in this segment, the speakers were not prompted by the interviewer, resulting in continuous spontaneous speech. This is expected to contain filled pauses, in line with Lallgee and Cook's (1969) study, which found that if a speaker feels pressure to keep speaking without having anything organised to say, they will fall back on filled pauses (Lallgee & Cook, 1969, p. 24).

[1] As judged by the researcher.

2.3 Annotation

All sound files were annotated using Praat (Boersma & Weenink, 2017). Four annotation tiers were used, detailing the language spoken, the type of speech, the position and type of filled pause and that of the vowel in the filled pause. The filled pauses were located by listening to the recordings, and the boundaries of each filled pause were added, aided by the waveform. The initial boundary of each filled pause was set at the vocalic onset.

2.4 Analysis

As the question the present study aims to answer is to whether Dutch speakers remain recognisable by their filled pauses when speaking English as opposed to Dutch, the filled pauses were analysed in each language according to a number of parameters. These are the number of filled pauses per minute, the number of vocalic and vocalic-nasal pauses, the duration of the filled pauses, and their first and second formants.

The first parameter, the number of filled pauses per minute, was obtained by dividing the total speaking time[2] per speaker by the number of filled pauses they used. The second, the number of vocalic and vocalic-nasal filled pauses, was obtained by tallying each type. Using a script in Praat, the duration (s) and first and second formants (Hz) of the filled pauses were extracted from the recordings for 19 speakers, as one speaker was excluded due to technical difficulties.

2.4.1 Analysis per language

First, mean speaking time and number of filled pauses in total for each language was calculated, followed by mean number and duration of each type of filled pause. In order to

[2] De Leeuw (2007) attained speaking time by subtracting the time the interviewer spoke and sections of silence greater than 2s from the total duration of the interview. As there was no interviewer interference in the recordings used in the present study, and the sections of silence were not so long as to be disruptive, the speaking time as used here was calculated from the moment the speaker started speaking to the moment he or she finished.

compare filled pauses per minute in English and Dutch, the means for each language were compared using a paired t-test in SPSS. A chi-squared test was used to determine the distribution of vocalic and vocalic-nasal filled pauses in the two languages. The mean duration, F1 and F2 of each type of filled pause were also compared between the two languages using a paired-test. For this analysis, the cases in which the standard deviation for F1 or F2 was more than 100 Hz were excluded, as they deviated too much from the mean, and would have skewed the results.

2.4.2 Analysis per speaker

The distribution of vocalic and vocalic-nasal filled pauses and filled pauses per minute for each speaker were determined using a chi-squared test. Before the duration and F1 and F2 of filled pauses were calculated for each speaker, the dataset was checked to establish whether there was enough data for each speaker to be able to run an accurate test of significance. Eight *uh*s or *um*s per speaker was deemed suffice to this end. Therefore, all speakers who did not have eight or more *uh*s or *um*s in both English and Dutch, as well as a standard deviation of less than 100 Hz for F1 or F2, were excluded from further analysis. This was done in order to exclude the outliers that might have occurred due to experimental errors. For each of the remaining 13 speakers, duration and F1 and F2 of their filled pauses were compared for English and Dutch, using a paired t-test.

3. Results

3.1 Results per language

Table 1 shows that when speaking English, the speakers spoke for longer than they did when speaking Dutch. The number of filled pauses per minute did not differ significantly for the two languages; $t(19) = 1.95$, $p=.066$.

Table 1.
Mean (m) and Standard Deviation (sd) of Speaking Time and Filled Pauses per Minute per Language across 20 Speakers

	Dutch		English	
	m	sd	m	sd
Speaking time (s)	131.99	20.62	141.63	15.82
FP per minute	9.64	4.90	8.03	3.43

As shown in table 2, speakers use *uh* more often than *um* in Dutch, and *um* more often than *uh* in English. This difference was significant; $x^2(1) = 17.2$, $p<.001$.

Table 2.
Number of Vocalic (uh) and Vocalic-Nasal (um) Filled Pauses per Language across 20 Speakers

	Uh	Um
Dutch	182	100
English	110	127

Table 3.
Mean (m) and Standard Deviation (sd) of Duration (s), F1 and F2 (Hz) per Language, and per type of Filled Pause

	Uh			Um		
	Duration	F1	F2	Duration	F1	F2
Dutch	0.296 (0.684)	581.07 (63.16)	1563.51 (128.39)	0.519 (0.189)	610.81 (72.15)	1555.67 (162.56)
English	0.314 (0.083)	621.60 (67.68)	1521.22 (126.22)	0.533 (0.182)	637.37 (64.99)	1495.62 (151.02)

There was no significant difference in duration of *uh* in Dutch versus English; $t(17) = 1.45$, $p=.167$. F1 for *uh* was significantly higher in English than it was in Dutch; $t(17) = 5.24$. $p<.001$, whereas F2 for *uh* was significantly higher in Dutch than in English; $t(17) = 3.46$, $p=.003$.

The mean duration of *um* did not differ significantly for the two languages; t(18) = 0.47, p=.644, just as the F1 for *um* did not; t(17) = 1.90, p=.074. F2 for *um* was significantly higher in Dutch than in English; t(17) = 3.38, p=.004.

3.2 Analysis per speaker

As shown in the previous section, the difference between English and Dutch for the parameters discussed was not significant in all cases. Therefore, this section will detail each of the four parameters for each speaker separately, in order to establish in regards to which parameters speakers use filled pauses similarly or differently in English and Dutch.

Figures 1, 2 and 3 (appendix A) show the number of filled pauses for each speaker, and the proportion of vocalic and vocalic-nasal filled pauses in both Dutch and English. There are no discernible trends in this data, as speakers differ greatly in the quantity of filled pauses they use and the varieties thereof, in both languages.

Tables 5, 6 and 7 (appendix B) show the duration, F1 and F2 of *uh* per speaker, and the significance of the differences between the two languages, respectively. Duration of *uh* was significantly different in the two languages for only one speaker, 1m, whose *uh* was longer in English than in Dutch. F1 was significantly different in English and Dutch for all speakers. Of these speakers, eight had a higher F1 in English, while two had a higher F2 in Dutch. F2 for *uh* was significantly higher in Dutch than in English for two speakers, whereas there was no difference for the other eight speakers.

Tables 8, 9 and 10 (appendix B) convey the duration, F1 and F2 of *um* and the significance of the difference between Dutch and English for five speakers. The duration of *um* was significantly different in Dutch and English for two speakers; one speaker, s001f1-1, realised a longer *um* in Dutch, whereas the other speaker, s016f1-1, realised a longer *um* in English. F1 was significantly higher in English for three of the speakers; one speaker had a

higher F1 in Dutch and the last speaker did not a have significantly different F1 for English or Dutch. F2 was higher in Dutch for three speakers and not significantly different for the other two speakers.

4. Discussion

The present study has explored the extent to which Dutch native speakers use filled pauses differently in Dutch and L2 English. This was done based on a number of aspects, namely: the number of filled pauses used by the speakers, the proportion of vocalic and vocalic-nasal filled pauses, their duration and their first and second formants. These aspects will be discussed here, first for English and Dutch in general, then for each speaker individually.

4.1 Comparison between languages

The results regarding filled pauses per minute in Dutch and English indicate that the speakers did not use more filled pauses per minute when speaking Dutch as opposed to English. This differs from results found by De Leeuw (2007), as she found that Dutch speakers used significantly more filled pauses than English speakers. These findings are, however, in line with results from Künzel's (1997) study, which posit that speakers transfer their hesitation behaviour from their native to their second language.

As table 2 shows, the speakers tended to use more vocalic than vocalic-nasal filled pauses when speaking Dutch. This is consistent with findings of Swerts et al.'s (1996) study, which found that, in Dutch, vocalic-nasal hesitation markers occurred less frequently than vocalic hesitation markers (as cited in De Leeuw, 2007, p. 107). When speaking English, the speakers in the present study used more vocalic-nasal than vocalic filled pauses. This is consistent with Lickley's (1994) study, which found that vocalic-nasals were more common for British English speakers (as cited in De Leeuw, 2007, p. 107).

While the speakers examined in the present study had different levels of proficiency in English, it was a second language for all of them. As Clark and Fox Tree (2002) found, speakers tend to use *uh* for short delays in speech, and *um* for longer ones (Clark & Fox Tree, 2002, p. 101). It is therefore plausible that the speakers in the present study use *um* more often than *uh* in L2 English because they anticipate longer delays in L2 English than in their native Dutch.

There is very little literature on the duration of filled pauses, and even less on the language-specificity of duration. A study by Goldman-Eisler (1961) found that, in English, filled pauses typically range between 0.2 and 0.8 seconds (Goldman-Eisler, 1961, p. 20). This study did not differentiate between the two types of filled pauses, however. The present study did differentiate between the two, but found no significant difference in duration of *uh* or *um* for either Dutch or English.

As the speakers examined in the present study had a varying proficiency in English, duration of (as well as number of) filled pauses was anticipated to be an indicator of proficiency, as filled pauses are used when a speaker encounters a problem during speaking. However, as proficiency was not measured and did not weigh in on the analysis, it is possible that this was a confounding factor.

F1 and F2 for *uh* differed significantly for the two languages: F1 was higher in English than in Dutch, and F2 was higher in Dutch than in English. This means that in L2 English, the [ə] in *uh* is a more open and back vowel, whereas in Dutch, it is more close and fronted. This is in line with what is described in Collins et al. (2011, p.58). This indicates that speakers realise *uh* as it is realised in their target language.

F1 for *um* was not significantly different in English and Dutch, whereas F2 was higher in Dutch than in English, signifying that the vowel in *um* is more fronted in Dutch, but that the vowel height is similar for the two languages. This indicates a degree of transfer from

the speakers' native language to their L2, as described by Künzel (1997).

The fact that there is transfer from native Dutch to L2 English for some of the aspects of filled pauses indicates that there might be some degree of speaker recognisability. As a result of Künzel's (1997) study, it was expected that there would be transfer of all aspects of filled pauses, including vowel quality, yet the results show that this is not the case. This means that the realisation of filled pauses in L2 English is not transferred in its entirety from native Dutch.

4.2 Comparison between languages, per speaker

The filled pauses per minute measure at the speaker-level shows a fair amount of speaker idiosyncrasy. The number of filled pauses per minute for both languages varies from, for example, 3.17 for speaker 6f to 17.09 for speaker 7m. As mentioned previously, the speakers' proficiency in English varied considerably, which could explain this difference to some extent. As can be seen in figure 1, however, for some of speakers, the number of filled pauses per minute is comparable for the two languages, signifying speaker-specificity, rather than an indicator of proficiency.

The majority of the speakers use the same proportion of *uh* and *um* in Dutch and in English, which is in line with Künzel's (1997) findings that speakers are quite consistent in using their respective personal variant of filled pauses. Some speakers, however, use either *uh* or *um* in a way that is consistent with findings from previous studies, as mentioned in section 4.1.2, and while this could be an indicator of proficiency, it could also mean that the speakers merely know how to use filled pauses in the two languages and employ this pausing strategy.

Examining duration of *uh* at the speaker-level shows that for all but one speaker, there is no significant difference in duration between Dutch and English. For all speakers, the difference in duration of *um* in Dutch versus English is significant in two cases. As mentioned

previously, duration could be an indicator of proficiency, but given the difference over these speakers, proper quantification of proficiency as a variable would be necessary in order to establish its effects.

F1 for *uh* is significantly higher in English for eight speakers, and higher in Dutch for two speakers. These results indicate that eight speakers realise [ə] in *uh* as a more open vowel in English, whereas two speakers realise [ə] as a more close vowel. F2 for *uh* is higher in Dutch than English for two speakers and not significantly different for the other eight. A higher F2 in Dutch indicates a more fronted vowel for *uh*. F1 for *um* is significantly different in the two languages for four speakers, of which 3 had a higher F1 in English and one had a higher F1 in Dutch, signifying a more open vowel. F2 for *um* differed significantly for three speakers, all of which had a higher F2 in Dutch. This indicates that the [ə] in *um* is more fronted.

For the majority of the speakers, F1 differed significantly for English and Dutch. However, the language for which F1 was higher was not consistent over the speakers. The same is true for F2. The fact that F1 differs in English and Dutch indicates varying degrees of vowel height, which suggests that speakers change the realisation of their filled pauses when speaking a second language, which would contradict Baldwin and French's (1990) transfer theory.

Of the four aspects of filled pauses examined in the present study, duration, type and number of filled pauses are aspects that are relatively simple to alter. If a speaker were to know how these aspects work in a target language, he or she could adopt them in order to sounds more native-like. Vowel quality is more difficult to alter, which makes it all the more interesting that the majority of the speakers had significantly different first formants for their filled pauses.

4.3 Limitations and further research

In light of all that has been described above, the biggest limitation of this study was failing to quantify the speakers' proficiency in English, as this has most likely been a confounding factor in much of the analysis. Further research on this subject should address this issue.

Further research might also benefit from determining whether or not speakers are aware of how filled pauses are realised in their target language. Formant frequency differed for the two languages examined in the present study, which indicates that speakers do not transfer this aspect of filled pauses from native to target languages, while other aspects are transferred. If speakers were to know how to realise filled pauses in the target language, there would be no need for transfer, resulting in significantly different outcomes in both languages, as was the case for formant frequencies. In order to measure this, however, further research would have to uncover whether speakers are aware of filled pauses in other languages implicitly, so as not to confound the results.

The present study used the amount of filled pauses per minute of speech for each speaker to quantify how many filled pauses were used, while the measure employed by Goldman-Eisler (1961) would likely have been more accurate. Goldman-Eisler calculated the percentage of filled pauses for the total number of words spoken, and while this method might have been a more accurate measure of the number of filled pauses used by the speakers for each language, it was not used in the present study due to time constraints.

5. Conclusion

The present study has aimed to determine the extent to which Dutch speakers remain recognisable by their use of filled pauses in native Dutch and L2 English. A number of parameters were used to quantify the difference in use for the two languages: filled pauses per minute, the proportion of vocalic and vocalic-nasal filled pauses, duration, F1 and F2 of filled

pauses. The statistical analysis of each of these aspects of filled pauses revealed that there was no significant difference for filled pauses per minute or duration for Dutch versus English, indicating transfer from speakers' native Dutch to L2 English. As for the proportion of vocalic and vocalic-nasal filled pauses, *uh* was used more in Dutch, and *um* was used more in English, which was consistent with previous research. The first and second formants are more problematic, as there was no clear trend, either between speakers or between languages. These results indicate a certain extent of transfer from native Dutch to L2 English, and thus a degree of speaker-recognisability.

References

Arnold, J. E., Fagnano, M., & Tanenhaus, M. K. (2003). Disfluencies signal theee, um, new information. *Journal of psycholinguistic research*, *32*(1), 25-36.

Boersma, P., & Weenink, D. (2017). Praat: doing phonetics by computer [Computer program]. Version 5.3.51, retrieved 6 November 2017 from http://www.praat.org/

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

Bosker, H. R., Quené, H., Sanders, T., & de Jong, N. H. (2014). Native 'um's elicit prediction of low-frequency referents, but non-native 'um's do not. *Journal of memory and language*, *75*, 104-116.

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

Broen, P. A., & Siegel, G. M. (1972). Variations in normal speech disfluencies. *Language and Speech*, *15*(3), 219-231.

Clark, H. H., & Wasow, T. (1998). Repeating words in spontaneous speech. *Cognitive psychology*, *37*(3), 201-242.

Clark, H. H., & Tree, J. E. F. (2002). Using uh and um in spontaneous speaking. *Cognition*, *84*(1), 73-111.

Collins, B., den Hollander, S. P., Mees, I. M., & Rodd, J. (2006). Sounding better. *A Practical Guide to English Pronunciation for Speakers of Dutch*.

De Leeuw, E. (2007). Hesitation markers in English, German, and Dutch. *Journal of Germanic Linguistics*, *19*(2), 85-114.

Fruehwald, J. (2016). Filled Pause Choice as a Sociolinguistic Variable. *University of Pennsylvania Working Papers in Linguistics*, *22*(2), 6.

Goldman-Eisler, F. (1961). A comparative study of two hesitation phenomena. *Language and Speech*, *4*(1), 18-26.

Jessen, M. (2008). Forensic phonetics. *Language and Linguistics Compass*, *2*(4), 671-711.

Künzel, H. F. (1997). Some general phonetic and forensic aspects of speaking tempo. *Forensic Linguistics*, *4*, 48-83.

Lallgee, M. G., & Cook, M. (1969). An experimental investigation of the function of filled pauses in speech. *Language and Speech*, *12*(1), 24-28.

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word*, *15*(1), 19-44.

Orr, R., Quené, H., Beek, R. V., Diefenbach, T., Leeuwen, D. A. V., & Huijbregts, M. (2011). An international English speech corpus for longitudinal study of accent development. In *Twelfth Annual Conference of the International Speech Communication Association*.

Swerts, M. (1998). Filled pauses as markers of discourse structure. *Journal of pragmatics*, *30*(4), 485-496.

Tree, J. E. F. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of memory and language*, *34*(6), 709-738.

Tree, J. E. F. (2001). Listeners' uses of um and uh in speech comprehension. *Memory & cognition*, *29*(2), 320-326.

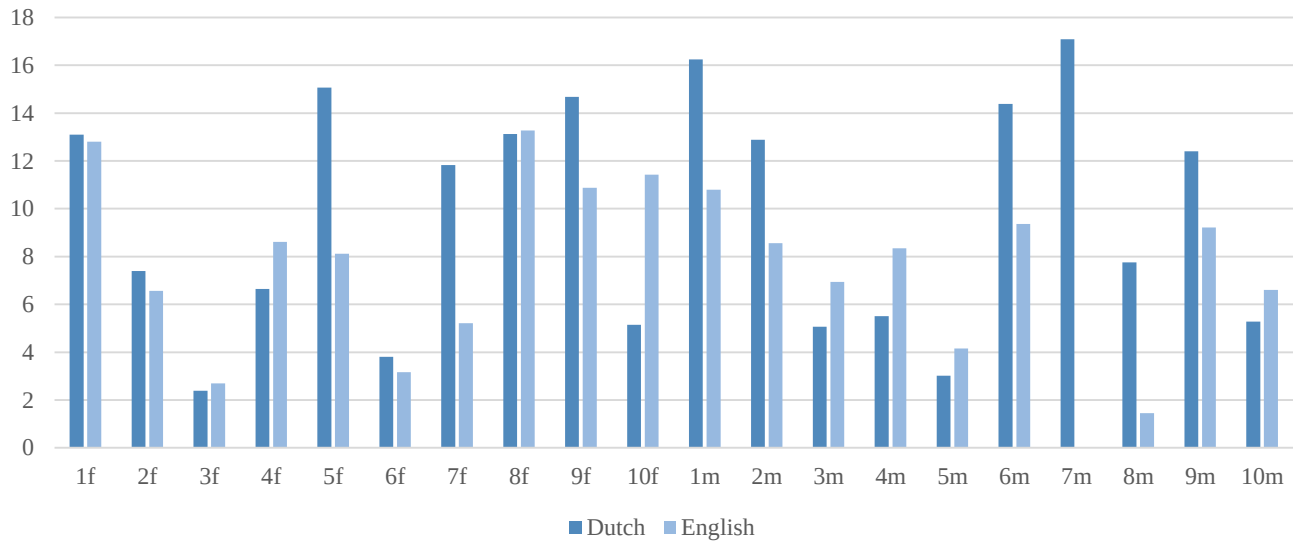
Tree, J. E. F. (2007). Folk notions of um and uh, you know, and like. *Text & Talk-An Interdisciplinary Journal of Language, Discourse Communication Studies*, *27*(3), 297-314.

Van Donzel, M. E., & Koopmans-van Beinum, F. J. (1996, October). Pausing strategies in discourse in Dutch. In *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on* (Vol. 2, pp. 1029-1032). IEEE.

Wieling, M., Grieve, J., Bouma, G., Fruehwald, J., Coleman, J., & Liberman, M. (2016). Variation and change in the use of hesitation markers in Germanic languages. *Language Dynamics and Change*, *6*(2), 199-234.

Appendix A

Figure 1.



Filled Pauses per Minute per Language, per Speaker

Figure 2.

Proportion of Uh *and* Um *per Speaker in Dutch*

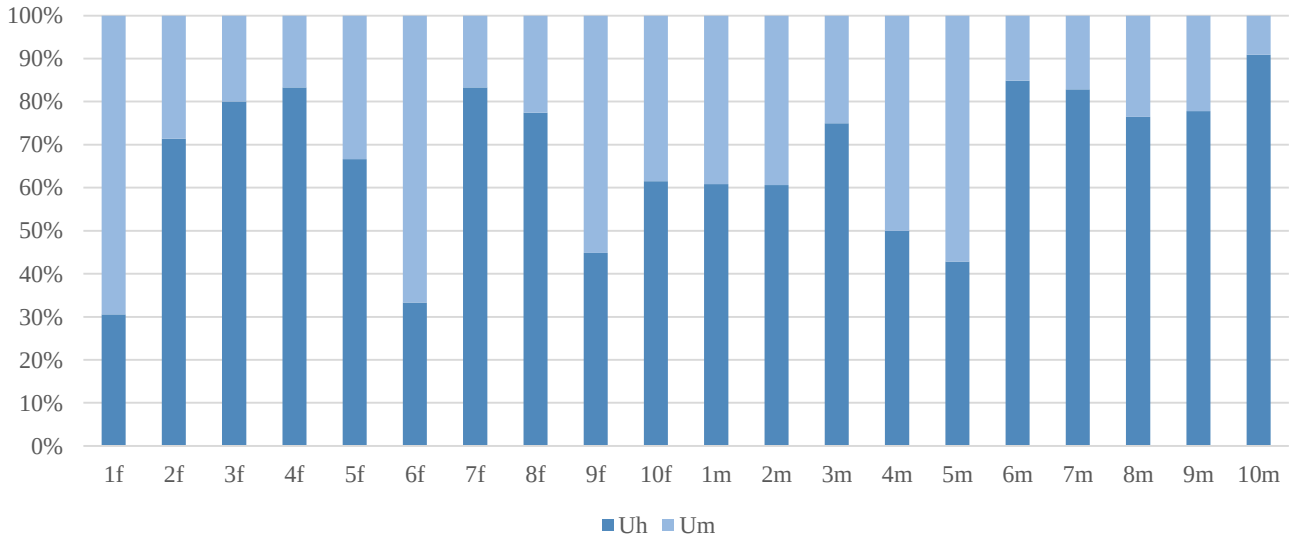
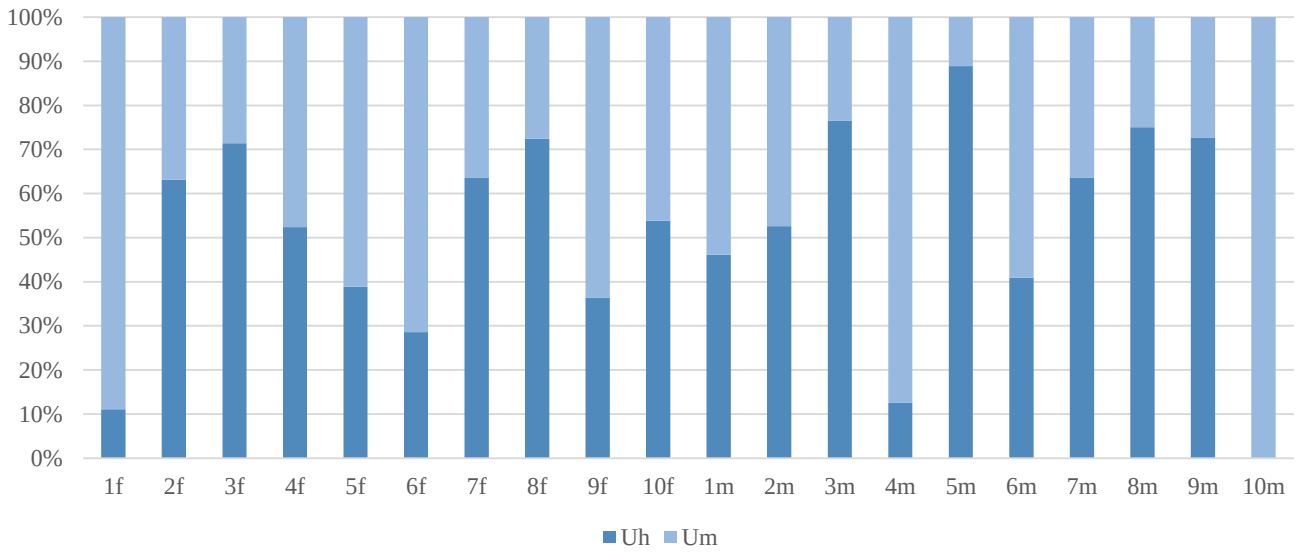


Figure 3.
Proportion of Uh *and* Um *per Speaker in English*



Appendix B

Table 5.
Mean (m) and Standard Deviation (sd) of Duration (s) of Uh *and Significance of Difference between Languages, per Speaker*

	Dutch		English		
	m	sd	m	sd	
2f	0.471	0.105	0.456	0.110	p=.744
4f	0.349	0.111	0.273	0.882	p=.104
8f	0.306	0.182	0.243	0.085	p=.215
10f	0.303	0.127	0.281	0.129	p=.776
1m	0.244	0.910	0.297	0.110	p=.048*
2m	0.230	0.072	0.280	0.117	p=.171
3m	0.279	0.167	0.362	0.178	p=.345
6m	0.208	0.104	0.310	0.126	p=.114
7m	0.303	0.145	0.342	0.152	p=.446
9m	0.308	0.137	0.343	0.170	p=.538

Table 6.

Mean (m) and Standard Deviation (sd) of F1 (Hz) of Uh *and Significance of Difference between Languages, per Speaker*

	Dutch		English		
	m	sd	m	sd	
2f	508.21	55.63	595.31	46.19	p<.001***
4f	666.39	37.94	653.62	40.59	p=.046*
8f	528.88	55.08	576.70	66.73	p<.001***
10f	668.23	59.02	640.57	47.29	p=.003**
1m	561.85	58.42	618.92	31.68	p<.001***
2m	476.97	25.78	530.26	82.23	p=.024*
3m	556.70	44.74	592.47	32.18	p=.004**
6m	493.11	41.28	650.60	36.98	p<.001***
7m	460.29	37.00	524.03	56.26	p<.001***
9m	526.75	57.83	563.75	37.72	p<.001***

Table 7.

Mean (m) and Standard Deviation (sd) of F2 (Hz) of Uh *and Significance of Difference between Languages, per Speaker*

	Dutch		English		
	m	sd	m	sd	
2f	1583.91	45.85	1542.06	37.52	p=.069
4f	1735.62	59.89	1697.98	73.62	p=.327
8f	1852.55	125.82	1821.01	162.93	p=.531
10f	1685.50	61.58	1621.53	151.73	p=.321
1m	1365.51	84.39	1304.83	66.04	p=.057
2m	1528.11	62.41	1490.91	132.33	p=.489
3m	1406.33	73.35	1352.52	60.61	p=.164
6m	1597.55	64.31	1431.58	75.71	p<.001***
7m	1479.13	72.71	1412.27	86.32	p=.027*
9m	1450.00	66.80	1472.03	51.24	p=.258

Table 8.

Mean (m) and Standard Deviation (sd) of Duration (s) of Um *and Significance of Difference between Languages, per Speaker*

	Dutch		English		
	m	sd	m	sd	
1f	0.427	0.070	0.376	0.200	p=.017*
5f	0.372	0.017	0.456	0.027	p=.014*
9f	0.359	0.111	0.346	0.107	p=.755
1m	0.440	0.117	0.458	0.081	p=.620
2m	0.385	0.130	0.449	0.137	p=.316

Table 9.

Mean (m) and Standard Deviation (sd) of F1 (Hz) of Um *and Significance of Difference between Languages, per Speaker*

	Dutch		English		
	m	sd	m	sd	
1f	671.55	9.700	704.29	12.01	p<.001***
5f	610.49	17.31	711.52	15.07	p<.001***
9f	718.22	95.09	776.87	51.49	p=.005**
1m	576.27	87.33	600.97	26.95	p=.256
2m	512.55	25.63	579.83	45.08	p<.001***

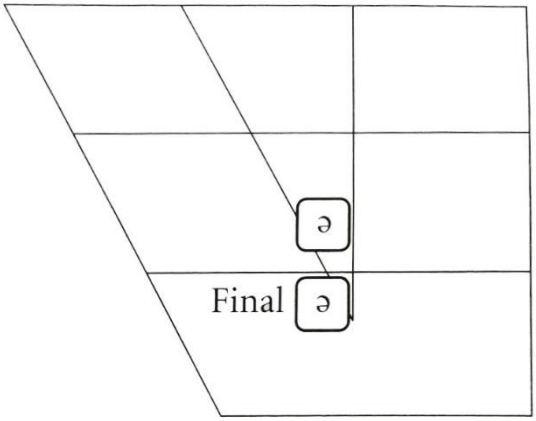
Table 10.

Mean (m) and Standard Deviation (sd) of F2 (Hz) of Um *and Significance of Difference between Languages, per Speaker*

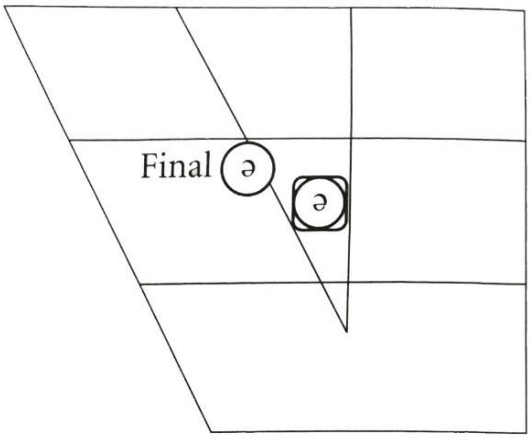
	Dutch		English		
	m	sd	m	sd	
1f	1482.26	13.36	1406.09	14.01	p<.001***
5f	1641.89	27.83	1483.46	25.21	p=.003**
9f	1526.59	77.94	1516.17	96.97	p=.779
1m	1360.85	45.73	1277.58	45.73	p=.021*
2m	1543.48	81.08	1509.39	116.05	p=.535

Appendix C

Figure 4.
Schwa vowel in English and Dutch, from Sounding Better: A Practical Guide to English Pronunciation for Speakers of Dutch *(Collins et al., 2006, p. 58).*



Schwa vowel in RP English



Schwa vowel in Dutch