

Nivja H de Jong*

Predicting pauses in L1 and L2 speech: the effects of utterance boundaries and word frequency

DOI 10.1515/iral-2016-9993

Abstract: This paper compares the distribution of silent and filled pauses in first (L1) and second language (L2) speech. The occurrence of pauses of 52 L2 and 18 L1 Dutch speakers was evaluated with respect to utterance boundaries and word frequency. We found that L2 speakers paused more often than L1 speakers within utterances; but not between utterances. Similarly, only within utterances, L2 pauses were longer than L1 pauses. Regarding word frequency, both L1 and L2 speakers are more likely to pause before lower frequency words as compared to higher frequency words. These findings imply that L1 and L2 speakers' production processes may be similar in that (1) pauses at utterance boundaries are used for conceptual planning mostly and (2) lexical retrieval difficulties are comparable for L1 and L2 speakers. These findings furthermore imply that when using fluency for L2 testing, pause locations must be taken into account.

Keywords: pause distribution, fluency, speech production, second language speaking

1 Introduction

Being fluent in an L2 is defined as being able to smoothly and effortlessly translate intended messages to speech (e.g., Schmidt 1992) and, hence, fluent speech is defined as speech without (unnatural) hesitations. Fluency, together with complexity (with respect to syntax and lexis) and accuracy in L2 speech (with respect to morpho-syntax, lexis, and pronunciation), is one of the three perceptual dimensions of speech that develop as L2 learners' proficiency progresses (Housen and Kuiken 2009). It is therefore no wonder that fluency in speech is used as a diagnostic in second language (L2) assessment. Indeed, human judgements on overall L2 proficiency are related to aspects of fluency (e.g., Iwashita et al. 2008). Additionally, for most speakers, their L2 speech is less fluent than their L1 speech and L2 speakers progress in certain aspects of fluency over time

*Corresponding author: Nivja H de Jong, Utrecht University, Trans 10, 3512JK Utrecht, Netherlands, E-mail: n.dejong@uu.nl

(Derwing et al. 2009; Towell et al. 1996). Aspects of fluency are indicative of proficiency and can be discriminatory between L1 and L2 speech. So could one simply count the number of disfluencies such as silent and filled pauses to distinguish between L1 and L2 speech or to measure a fluency aspect of L2 proficiency? The answer is no, because between L1 speakers there are also differences in levels of fluency (Bortfeld et al. 2001) and between L2 speakers matched on overall L2 proficiency, large differences with respect to measures of fluency likewise exist.

Davies (2003) suggested that it might not be the amount of disfluencies, but rather their distribution, that will be indicative of L2 proficiency. Davies (2003) refers to Pawley and Syder (1983), who suggest that native speakers produce speech “one clause at a time”. Because native speakers can produce multi-word chunks, they direct attention to planning the upcoming message, without having to pay attention to the linguistic formulation of the current message. This will result in pauses that occur at clause boundaries mostly. L2 speakers, however, may not (yet) have a large repository of pre-fabricated chunks and therefore pauses may also occur clause-medially.

To investigate how L2 speakers’ pause distribution is different from L1 speakers’ pause distribution, one first needs to consider the distribution of pauses for L1 speakers. The current study will investigate two aspects that have been shown to influence pause occurrence in L1 speech, and extend it to L2 speech: (1) pause occurrence between and within utterances, and (2) pause occurrence before low and high frequency words. In L1 speech, pauses have been found to be more likely to occur before (major) constituents, at syntactic boundaries (Swerts 1998), presumably reflecting pauses in which speakers plan what to say. Within clauses, pauses are more likely to occur before open-class words (Maclay and Osgood 1959), that are low in predictability (Goldman-Eisler 1958) and (therefore) pauses are more likely to occur before low-frequency words (Hartsuiker and Notebaert 2010; Kircher et al. 2004).

Previous research (e.g., Riazantseva 2001; Skehan and Foster 2007) has already investigated the potential difference between L1 and L2 pause distribution between versus within ASU’s (Analysis of Speech Units, Foster et al. 2000) and (major) constituents, and investigated how L2 proficiency moderates pause distributions. As such, this study will partly be a replication study. However, as argued below, the present study adopts better-suited analyses to investigate these questions. The factor “word frequency” has (to my knowledge) not been investigated in L2 speech before.

Additionally, the current study will investigate potential differences in pause durations between L1 and L2 speech, taking pause placement into account. Riazantseva (2001) also investigated differences in pause duration

between L1 and L2 speech and between speakers with different levels of proficiency. She found that higher proficient speakers paused, on average, longer than lower proficient speakers. Other studies, however, have suggested that pause duration is not related to proficiency (De Jong et al. 2013; De Jong et al. 2015; Towell et al. 1996). However, with the exception of De Jong et al. (2015), these studies did not take pause placement into account. The current study will therefore test whether pause duration is dependent on pause placement (between versus within ASU's) for L1 and L2 speakers.

1.1 Previous studies investigating the distribution of pauses

From previous research one can conclude that indeed, L2 speakers pause more often within ASU's (Skehan and Foster 2007), clauses (Tavakoli 2011) or constituents (Riazantseva 2001) than L1 speakers do. However, the measures and analyses that were used in these studies make firm conclusions difficult. Riazantseva (2001) used percentage of pauses within constituents and compared this measure for the same speakers in their L1 (Russian) and L2 (English). It is likely, however, that the speakers produced more complex and longer constituents in their L1 compared to their L2. If this is indeed the case, there were more opportunities to pause in the longer constituents (L1) than in the shorter constituents (L2). The found difference in percentage of pauses within constituents between L1 and L2 might therefore have been underestimated. Rianzantseva (2001) did not find a difference between the two L2 proficiency levels in her sample (higher and intermediate). Because it may very well have been the case that the higher proficient L2 speakers produced longer constituents, comparing the percentage of pauses within constituents between these groups may not have been valid.

The studies by Tavakoli (2011) and Skehan and Foster (2007) have the same issue with the measures used to compare the distribution of pauses between L1 and L2 speakers. Skehan and Foster (2007) report a *ratio* of number of pauses within ASU's to number of pauses between ASU's. Again, when longer ASU's are produced, there are more opportunities to pause within them than when learners produce only short ASU's. In addition, they report the mean number of pauses either within or between ASU's for the speaking performances. From these measures, it is found that L1 speakers pause *more* at the boundaries than L2 speakers do. However, if L1 speakers in total produce more speech and more ASU's, comparing the number of pauses for L1 and L2 speakers at this position is not a valid comparison.

Finally, Tavakoli (2011) counted the mean number of pauses either within or between clauses produced by L1 speakers and compared these to the mean

number of pauses for a group of L2 speakers. She also reported that L2 speakers pause more often within clauses than L1 speakers do. However, as explained before, when clauses are longer, which is likely the case for the L1 speakers, there is more opportunity to pause within them. In the current study, we will therefore employ logistic regression analyses, where each word boundary is considered as potential pause position. With such analyses, length of ASU's or number of ASU's in the sample are no longer confounding factors.

1.2 Research questions

The previous discussion leads to the following four research questions:

- RQ1: Do L1 and L2 speakers differ in pause distribution with respect to within/between ASU's?
- RQ2: Do L1 and L2 speakers differ in pause distribution with respect to word frequency?
- RQ3: What is the relation between pause placement (between/within ASU's) and pause duration for L1 and L2 speakers?
- RQ4: Does L2 proficiency moderate the pause occurrence and pause duration patterns of L2 speakers?

2 Method

The current study used the same L2 data as reported on in De Jong et al. (2015), but now adding transcripts and measures from L1 speakers performing the same speaking tasks.

2.1 Participants

Twenty-nine native speakers of English, 25 native speakers of Turkish, and 18 native speakers of Dutch were paid to take part in our experiment, with signed informed consent. This research was part of a larger project with, for some participants, more tasks than are reported on here (see De Jong et al. 2012, 2013, and 2015). Depending on how many tasks the participants completed, they were paid between 30 and 50 euros. The Turkish (8 male, 17 female; mean age = 32, range = 23–48) and English participants in the current study (11 male, 18 female; mean age = 31, range = 23–43) had come to the Netherlands between the ages of 18 and 40 (English range = 22–40, Turkish range = 18–35). Most of these participants had lived in the Netherlands for

fewer than 10 years (English mean = 4.5 years, range = 1 month to 21 years; Turkish mean = 7 years, range = 9 months to 20 years). All L2 participants were at an intermediate to advanced level of Dutch as an L2 and the majority was taking intermediate or advanced level Dutch courses to prepare for enrollment at the University of Amsterdam. Most of the 18 native speakers (7 male, 11 female); mean age = 26, range = 19 – 45) were students at the same university.

2.2 Materials: speaking tasks

Speech was elicited by using eight speaking tasks as described in De Jong et al. (2012). The speaking tasks differed in difficulty, formality, and discourse mode. Instructions for each task contained specific information about the speaking task itself, which was provided by one or several visual-verbal cues on a computer screen. No additional knowledge about the topic beyond the information provided in the tasks was needed to successfully complete each speaking task. Information about the purpose and audience of the task was also provided.

2.3 Materials: vocabulary test

As a proxy for overall L2 proficiency, to be assessed separately from the speaking performances, we chose to use a productive vocabulary task. Vocabulary knowledge has been shown to be a good predictor of overall L2 proficiency (Beglar and Hunt 1999; Zareva et al. 2005). Moreover, the vocabulary test currently used has been shown to be a strong predictor of overall speaking proficiency. In De Jong et al. (2012), using structural equation modeling, the vocabulary knowledge score showed a strong relation to ratings of overall speaking proficiency ($r = 0.79$).

The paper-and-pencil task (with instructions in the L2, Dutch) elicited knowledge of 90 words, and of 26 multi-word units. The total score for each participant was calculated as the total number of correct responses. The scoring procedure was lenient towards spelling mistakes and errors in inflectional variants. This vocabulary test was piloted and improved before first used in De Jong et al. (2012). In their original sample of 181 L2 learners of Dutch, Cronbach's alpha for this test was high ($\alpha = 0.98$), indicating good reliability.

2.4 Procedure

Participants completed the eight tasks in an office with a native Dutch-speaking experimenter present. The tasks were presented on a computer screen and the

participants' speech was recorded. Participants navigated the experiment and instructions themselves. Each task consisted of screens containing information about the task in Dutch, including pictures. For each task, participants had 30 seconds of preparation time and 120 seconds of speaking time, which was shown by a status bar at the bottom of the screen. Participants could press a "finished" button if they finished the task before the 120 seconds had gone by.

The average time used for completing all eight tasks was around 25 minutes. The vocabulary task was done in a separate session and took between twenty minutes and an hour.

2.5 Data analyses

All speech recordings (72 participants, eight tasks; totaling roughly 15 hours of speaking data) were transcribed and annotated by three native speakers of Dutch. They followed precise guidelines to annotate as similarly as possible. For two (English) participants, more than half of the performances were not recorded well (interference with the computer caused a strong hum). Therefore, these two participants were discarded from further analyses. In addition, eight recordings (from 6 participants) were not recorded with sufficient quality to make precise transcriptions. These recordings were also discarded. From 70 participants we thus obtained speech performances from at least six, but for most participants eight tasks. The transcriptions were made in CLAN (MacWhinney 2000). Besides orthographic transcriptions, information relevant for measuring pauses was inserted. Silent pauses were detected by careful listening and by using the waveform (as shown in CLAN). The transcribers added silent pause boundaries in the CLAN-waveform manually. The silent pauses were subsequently measured in milliseconds automatically (using the boundaries). The lower silent pause threshold was set at 250 ms, to exclude short so-called micropauses (Riggenbach 1991), which are irrelevant for measures of L2 fluency (De Jong and Bosker 2013).

The transcriptions were also split up into so-called analysis of speech units (ASU). Foster et al. (2000) have shown that using the ASU is the optimal way of dividing transcribed data into analyzable units. As defined by Foster et al. (2000: 365), an ASU is "a single speaker's utterance consisting of an independent clause, or a subclausal unit, together with any subordinate clause(s) associated with either." In the current analyses, we will use such ASU's (the actual speech units) to mark major boundaries in speech. Silent pauses were categorized as being either between or within ASU. Furthermore, the transcripts were annotated with nonlexical filled pauses (such as "uh," "uhm," "er," "mm"). Finally, for nouns, CELEX word frequency was added (Baayen et al. 1995).

3 Results

In total, from 70 participants, 84599 words were transcribed, 1209 on average per participant (SD 369) with a range from 380 to 2013. Below, we report on a number of analyses. We ran generalized linear mixed effect models as implemented in the lme4 library (Bates et al. 2012) in R to predict either silent or filled pause occurrence separately. With such analyses, we were able to consider each word transition as a data point while accounting for variance between speakers and variance between words. Taking this variance into account is important, as for both speakers and words we have repeated measures. Each word transition can be considered a data point because between two words either a pause occurred, or did not occur. For both silent and filled pauses, we first ran models to test for differences with respect to ASU-position (between or within ASU). Taking only the transitions within ASU's into account, for all transitions to nouns, a potential effect of Word Frequency was tested. To test interactions with Proficiency (measured with Vocabulary knowledge) and MotherTongue (English/Turkish), analyses were run on the subset of L2 speakers.

Finally, we ran linear mixed models to predict silent pause duration. Again, we ran models investigating the relation between pause placement and pause duration with all participants to test interactions with L1/L2 and separate models on the subset of L2 speakers, to test interactions with Proficiency and MotherTongue. No analyses to test for the effect of word frequency on pause duration for the subset of nouns was viable, as this restriction led to a too small dataset.

3.1 Predicting pause occurrence: between or within ASU's

Table 1 shows the counts of silent and filled pauses either between or within ASU's for both L1 and L2 speech. The percentages are based on the total number

Table 1: Number of silent and filled pauses (percentages*) for L1 and L2 speakers.

	Nr of silent pauses (% of total)		Nr of filled pauses (% of total)	
	L1	L2	L1	L2
Between ASU	1320 (73.5 %)	3805 (74.4 %)	175 (9.7 %)	533 (10.4 %)
Within ASU	1599 (7.7 %)	7572 (15.1 %)	684 (3.3 %)	4257 (8.5 %)

Note: *Percentages based on total number of ASU's for L1 and L2 (1797 and 5115) and total within ASU-words for L1 and L2 (20660 and 50115).

of ASU's for L1 and L2 speakers (1797 and 5115, respectively) and on the total number of words (minus all the initial words of ASU's) produced for L1 and L2 speakers (20660 and 50115, respectively). From these numbers, we can already see that L1 speakers, on average, had longer ASU's than L2 speakers (13.5 and 11.8 words per ASU, respectively; $\chi^2 (1) = 26.12$, $p < 0.001$).

To test for differences between L1 and L2 speech with respect to pause occurrence, we first ran generalized linear mixed models predicting silent pause occurrence for each word with participant ($N = 70$) and word ($N = 4110$) as crossed random effects, using the Laplace approximation. Both random effects were highly significant, as shown by likelihood ratio tests (Pinheiro and Bates 2000). We then proceeded with adding fixed effects; again using likelihood ratio tests to test whether adding these fixed effects significantly improved the model. We added Task (tasks 1 – 8), Position (within or between ASU), and NativeSpeaker (L1 or L2) as fixed effects. Task as effect was not to answer a research question, but was added to get more precise estimations of the remaining effects (the effects of the tasks are therefore mentioned in the tables presenting the models, but will not be commented on). Finally, to test whether a possible effect for Position was similar across L1 and L2 speakers, we added the interaction between Position and NativeSpeaker. The first three columns of Table 2 show the resulting model. This model took L2 speaker in task 1

Table 2: Results of generalized linear mixed models predicting silent pause occurrence and filled pause occurrence for L1 and L2 speakers' word transitions.

	Predicting silent pause occurrence			Predicting filled pause occurrence		
	Estimates (SE)	z-values	p-values	Estimates (SE)	z-values	p-values
<i>Fixed effects</i>						
(Intercept)	0.742 (0.086)	8.63	< 0.001	-1.960 (0.131)	-14.95	< 0.001
Nativespeaker	-0.098 (0.144)	-0.68	0.495	-0.011 (0.234)	-0.05	0.961
ASU within	-2.659 (0.042)	-63.27	< 0.001	-0.674 (0.059)	-11.52	< 0.001
Nativespeaker with ASU within	-0.719 (0.074)	-9.75	< 0.001	-0.856 (0.106)	-8.09	< 0.001
Task2	0.153 (0.053)	2.89	0.004	0.028 (0.069)	0.41	0.683
Task3	-0.094 (0.047)	-1.99	0.047	-0.312 (0.065)	-4.83	< 0.001

(continued)

Table 2: (continued)

	Predicting silent pause occurrence			Predicting filled pause occurrence		
	Estimates (SE)	z-values	p-values	Estimates (SE)	z-values	p-values
Task4	0.027 (0.049)	0.55	0.581	-0.042 (0.064)	-0.65	0.516
Task5	0.187 (0.048)	3.88	< 0.001	-0.044 (0.064)	-0.68	0.498
Task6	0.165 (0.046)	3.56	< 0.001	0.248 (0.059)	4.22	< 0.001
Task7	-0.136 (0.047)	-2.91	< 0.001	-0.153 (0.062)	-2.49	0.013
Task8	-0.015 (0.045)	-0.33	0.745	-0.039 (0.060)	-0.65	0.516
<i>Random effects</i>						
Words	0.511			0.410		
Speakers	0.217			0.603		

with position between ASU's as its intercept, and all effects should be interpreted relative to this intercept. Between ASU, there was no effect for NativeSpeaker. There was an effect of Position: for each word transition, there were lower probabilities of pauses to occur when the word would be within ASU's as compared to between ASU's. Crucially, there was also a significant interaction between NativeSpeaker and Position showing that the effect of Position (fewer pauses within than between ASU's) is stronger for L1 speech than for L2 speech.

For filled pauses, we ran the same model, but now predicting filled pause occurrence. The results of this model are shown in the last three columns of Table 2. The results were very similar to the model for silent pauses: no significant effect for NativeSpeaker between ASU's, a significant effect of Position (lower chance of pauses within ASU's), and a significant interaction: the effect of Position was again stronger for L1 speech as compared to L2 speech.

We then tested, for the subset L2 speakers, whether there was an effect of Proficiency (estimated by the vocabulary size measure) and MotherTongue (Turkish or English). This model predicted silent pause occurrence for each word with L2 participant ($N = 52$) and word ($N = 3350$) as crossed random effects, using the Laplace approximation. The first three columns of Table 3 show the result of this model, which took English speakers' pausing behavior between ASU's in task 1 as intercept. Relative to this intercept, there was no overall effect of MotherTongue. There was an effect of Proficiency (higher Proficiency resulted in a lower probability of pausing between ASU's) and an effect of Position

Table 3: Results of generalized linear mixed models predicting silent pause occurrence and filled pause occurrence for L2 speakers' word transitions.

	Predicting silent pause occurrence			Predicting filled pause occurrence		
	Estimates (SE)	z-values	p-values	Estimates (SE)	z-values	p-values
<i>Fixed effects</i>						
(Intercept)	0.645 (0.102)	6.34	< 0.001	-2.174 (0.178)	-12.23	< 0.001
Proficiency	-0.139 (0.063)	-2.21	0.027	-0.239 (0.115)	-2.08	0.037
L1Turkish	0.182 (0.128)	1.42	0.157	0.280 (0.238)	1.18	0.238
ASU within	-2.632 (0.055)	-47.91	< 0.001	-0.813 (0.081)	-10.04	< 0.001
Proficiency with ASU within	-0.164 (0.037)	-4.44	< 0.001	0.026 (0.048)	0.54	0.590
L1Turkish with ASU within	-0.126 (0.074)	-1.71	0.087	0.303 (0.103)	2.95	0.003
Task2	0.157 (0.059)	2.66	0.008	0.007 (0.076)	0.10	0.923
Task3	-0.043 (0.053)	-0.80	0.422	-0.292 (0.070)	-4.15	< 0.001
Task4	0.070 (0.054)	1.28	0.199	-0.036 (0.070)	-0.52	0.606
Task5	0.243 (0.054)	4.53	< 0.001	-0.017 (0.070)	-0.25	0.804
Task6	0.201 (0.053)	3.82	< 0.001	0.227 (0.066)	3.47	< 0.001
Task7	-0.131 (0.053)	-2.48	0.013	-0.079 (0.067)	-1.18	0.240
Task8	-0.050 (0.052)	-0.98	0.329	0.003 (0.065)	0.04	0.968
<i>Random effects</i>						
Words	0.518			0.433		
Speakers	0.153			0.602		

(within ASU's there was a lower probability of a silent pause than between ASU's). The interaction between Proficiency and Position was also significant: for participants with a higher Proficiency, there was a stronger effect of Position (higher probability of pausing within ASU's as compared to between ASU's). There was no interaction between MotherTongue and Proficiency, indicating that Turkish and English native speakers do not show differential pause probabilities in their L2 (Dutch).

For filled pauses, we tested the same model, but now had filled pause occurrence as the dependent variable. This model is shown in the last three columns of Table 3. Similar to the model for silent pauses, there was no significant effect for Mothertongue. The significant interaction between Mothertongue and Position showed, however, that within ASU's, the Turkish speakers used more filled pauses than the English speakers. There was again an effect of Proficiency: higher proficiency was associated with fewer filled pauses. Because there was no interaction between Position and Proficiency, we can conclude that this effect of Proficiency was similar between and within ASU's.

3.2 Predicting pauses: word frequency

For this analysis, we decided to focus on content words and restrict the analysis to nouns only. We excluded adjectives and verbs as it is unclear which frequency measure would actually be relevant due to the different inflectional variants of the adjectival and verb forms. In our corpus of speech, 14431 words were classified as nouns (within ASU's), 10644 tokens by L2 speakers, and 3787 tokens by L1 speakers. The range, median, and mean CELEX noun frequency was slightly lower for L1 compared to L2 speakers when calculated over all tokens (L1: range 0 – 55630, median: 2518, mean: 8098; L2: range 0 – 55630, median: 3112, mean: 8536). Comparing the mean frequencies for L1 and L2 in a linear mixed model with NativeSpeaker as a predictor variable (and participant as random effect) indeed showed a significant difference between the frequencies of the nouns used by the two groups ($p = 0.038$).

Similarly as for the analysis on all word transitions, for the 14431 transitions to nouns both participant ($N = 70$) and word (i. e., nominal word, $N = 1221$) were highly significant random factors, both in the model predicting silent pauses (first three columns of Table 4) as for the model predicting filled pauses (last three columns of Table 4). Likewise for both models, the effect of NativeSpeaker was significant: before nouns L1 speakers are less likely to pause than L2 speakers. Also the factor Word Frequency proved to significantly improve both models; transitions to high-frequent nouns are less likely to contain a pause than transitions to low-frequent nouns. The interactions between NativeSpeaker and Word Frequency, however, were not significant.

To see whether Mothertongue and Proficiency were significant predictors for pauses occurring before nouns, we also ran generalized linear mixed models on the subset of L2 speaker data of all nouns ($N = 10644$). Again, Participant ($N = 52$) and Word (nominal word; $N = 988$) were the crossed

Table 4: Results of generalized linear mixed models predicting silent pause occurrence and filled pause occurrence for L1 and L2 speakers' transitions to nouns.

	Predicting silent pause occurrence			Predicting filled pause occurrence		
	Estimates (SE)	z-values	p-values	Estimates (SE)	z-values	p-values
<i>Fixed effects</i>						
(Intercept)	-2.271 (0.146)	-15.51	< 0.001	-2.845 (0.175)	-16.22	< 0.001
Nativespeaker	-1.064 (0.187)	-5.69	< 0.001	-0.856 (0.248)	-3.45	0.001
Noun frequency	-0.289 (0.053)	-5.41	< 0.001	-0.361 (0.054)	-6.65	< 0.001
Nativespeaker with Noun frequency	-0.157 (0.083)	-1.88	0.060	0.157 (0.095)	1.64	0.101
Task2	-0.400 (0.194)	-2.07	0.039	-0.673 (0.208)	-3.24	0.001
Task3	-0.566 (0.183)	-3.10	0.002	-0.442 (0.198)	-2.24	0.025
Task4	-0.243 (0.174)	-1.40	0.162	-0.105 (0.187)	-0.56	0.572
Task5	-0.088 (0.195)	-0.45	0.652	0.044 (0.209)	0.21	0.832
Task6	0.280 (0.153)	1.83	0.067	0.302 (0.164)	1.84	0.067
Task7	-0.219 (0.159)	-1.37	0.170	-0.415 (0.176)	-2.36	0.018
Task8	-0.148 (0.160)	-0.93	0.354	-0.053 (0.177)	-0.30	0.763
<i>Random effects</i>						
Words	0.605			0.525		
Speakers	0.326			0.631		

random effects. For silent pauses (see columns 1 – 3 of Table 5), the effects of Mother tongue, Proficiency, and Noun Frequency were significant: Turkish speakers produced more silent pauses before nouns than English speakers did; higher proficient speakers produced fewer silent pauses than lower proficient speakers; and there was a higher probability of silent pauses before lower-frequency words. There were no significant interactions. In the model predicting filled pauses, the results showed that the effect of Proficiency was not significant. Turkish speakers were more likely to use a filled pause than English speakers. Like in the model for silent pauses, there were no significant interactions.

Table 5: Results of generalized linear mixed models predicting silent pause occurrence and filled pause occurrence for L2 speakers' transitions to nouns.

	Predicting silent pause occurrence			Predicting filled pause occurrence		
	Estimates (SE)	z-values	p-values	Estimates (SE)	z-values	p-values
<i>Fixed effects</i>						
(Intercept)	-2.473 (0.163)	-15.18	< 0.001	-3.225 (0.218)	-14.76	< 0.001
Proficiency	-0.246 (0.071)	-3.45	0.001	-0.193 (0.113)	-1.70	0.088
L1Turkish	0.386 (0.145)	2.64	0.008	0.591 (0.236)	2.50	0.012
Noun frequency	-0.262 (0.064)	-4.09	< 0.001	-0.348 (0.067)	-5.17	< 0.001
Proficiency with Noun frequency	0.033 (0.033)	0.99	0.320	-0.032 (0.035)	-0.89	0.373
L1 Turkish with Noun frequency	-0.058 (0.068)	-1.00	0.315	-0.005 (0.077)	-0.06	0.952
Task2	-0.339 (0.206)	-1.65	0.099	-0.630 (0.222)	-2.84	0.005
Task3	-0.562 (0.194)	-2.90	0.004	-0.406 (0.211)	-1.92	0.055
Task4	-0.263 (0.187)	-1.41	0.160	-0.149 (0.203)	-0.73	0.463
Task5	-0.054 (0.208)	-0.26	0.796	0.019 (0.223)	0.09	0.930
Task6	0.265 (0.167)	1.59	0.112	0.352 (0.179)	1.96	0.049
Task7	-0.211 (0.172)	-1.23	0.220	-0.307 (0.188)	-1.64	0.101
Task8	-0.219 (0.174)	-1.26	0.207	-0.004 (0.191)	-0.02	0.982
<i>Random effects</i>						
Words	0.607			0.490		
Speakers	0.203			0.590		

3.3. Predicting silent pause duration: between or within ASU's

To further investigate differences in fluency between L1 and L2 speakers, we also analysed silent pause duration, taking into account pause location. Table 6 shows the means and standard deviations of all silent pause durations (of all silent pauses >250 ms and <3,000 ms) between and within ASU's for L1 and L2 speech separately.

Table 6: Means and standard deviations of silent pauses for L1 and L2 speech in milliseconds.

	Mean duration of silent pauses (SD)	
	L1	L2
Between ASU (N = 4995)	729 (407)	794 (464)
Within ASU (N = 9075)	565 (329)	687 (439)

In the analyses predicting silent pause duration, we used linear mixed models. We had Participant and Word as crossed random effects, as in the analyses predicting pause occurrence. We first investigated whether (log) pause duration in our dataset was dependent on Position (within or between ASU) and NativeSpeaker. There were 14070 silent pauses in the dataset. Both random effects Participant (N=70) and Word (N=1484) were significant. In addition, the fixed effect Task proved to be a significant factor. Table 7 shows the results of this model. The number of degrees of freedom required for statistical significance testing of the *t* values was calculated as $J-m-1$ (Hox 2010), where *J* is

Table 7: Results of linear mixed model predicting (log) silent pause duration for L1 and L2 speakers.

	Predicting (log) silent pause duration		
	Estimates (SE)	t-values	p-values
<i>Fixed effects</i>			
(Intercept)	6.627 (0.032)	205.91	< 0.001
Nativespeaker	-0.084 (0.056)	-1.50	0.139
ASU within	-0.181 (0.012)	-14.51	< 0.001
Nativespeaker with ASU within	-0.071 (0.023)	-3.04	0.004
Task2	-0.021 (0.020)	-1.06	0.294
Task3	-0.036 (0.019)	-1.92	0.060
Task4	-0.084 (0.020)	-4.28	< 0.001

(continued)

Table 7: (continued)

	Predicting (log) silent pause duration		
	Estimates (SE)	t-values	p-values
Task5	-0.041 (0.019)	-2.17	0.034
Task6	-0.055 (0.018)	-3.03	0.004
Task7	-0.093 (0.019)	-4.91	< 0.001
Task8	-0.090 (0.018)	-5.04	< 0.001
<i>Random effects</i>			
Words	0.0058		
Speakers	0.0275		
Residual	0.2315		

the most conservative number of a random effect (70 speakers) and m is the total number of explanatory variables in the model ($m = 13$) resulting in 56 degrees of freedom. The model took L2 speaker with position between ASU's as its intercept, and the effects should be interpreted relative to this intercept. Between ASU's, there was no effect for NativeSpeaker. There was an effect of Position: within ASU's pauses were shorter as compared to between ASU's. Crucially, there was also a significant interaction between NativeSpeaker and Position: the effect of Position (shorter pauses within than between ASU's) proved to be stronger for L1 speech than for L2 speech.

We then tested, for the subset of L2 speakers, whether there was an effect of Proficiency and Mothertongue. This model predicted silent pause duration for each silent pause in this subset with participant ($N = 52$) and Word ($N = 1332$) as crossed random effects and with $df = 36$ (Hox 2010). Table 8 shows the results of this model, which had English speakers between ASU's as intercept. Relative to this intercept, there was no effect of Mothertongue or Proficiency, which means that between ASU's, Turkish and English speakers had comparable pause durations, no matter what their Proficiency was. The main effect for Position shows that the pauses within ASU's were significantly shorter than pauses between ASU's. Additionally, there were two significant interactions. First of all, the effect of Position was modulated by Proficiency: the higher the Proficiency of the L2 speakers, the shorter the pause durations within ASU's tended to be. Secondly, it turned out that within ASU's, the Turkish speakers tended to have shorter pause durations than the English speakers.

Table 8: Results of linear mixed model predicting (log) silent pause duration for L2 speakers.

	Predicting (log) silent pause duration		
	Estimates (SE)	t-values	p-values
<i>Fixed effects</i>			
(Intercept)	6.523 (0.038)	172.53	< 0.001
Proficiency	0.040 (0.025)	1.65	0.108
L1Turkish	0.060 (0.050)	1.22	0.230
ASU within	-0.121 (0.015)	-7.85	< 0.001
Proficiency with ASU within	-0.052 (0.010)	-5.05	< 0.001
L1Turkish with ASU within	-0.043 (0.020)	-2.09	0.044
Task2	-0.014 (0.021)	-0.65	0.520
Task3	-0.037 (0.020)	-1.89	0.067
Task4	-0.050 (0.020)	-2.45	0.019
Task5	-0.026 (0.020)	-1.32	0.195
Task6	-0.042 (0.019)	-2.20	0.034
Task7	-0.067 (0.020)	-3.42	0.002
Task8	-0.074 (0.019)	-3.95	< 0.001
<i>Random effects</i>			
Words	0.006		
Speakers	0.028		
Residual	0.243		

4. Discussion and conclusion

This paper examined how L1 and L2 speech differs in pause placement regarding utterance boundaries and word frequency. Additionally, the paper investigated whether L2 proficiency is a mediating factor in predicting pause distribution. Finally, silent pause duration between and within utterances was investigated.

Previous studies had already shown that L2 speakers tend to pause more within constituents, clauses, and ASU's as compared to L1 speakers (Riazantseva 2001; Skehan and Foster 2007; Tavakoli 2011). The measures and analyses used in these studies, however, potentially underestimated such differences. The current study therefore used logistic analyses, taking all word transitions as potential pause locations to predict, for each potential pause location, whether a pause actually occurred or not.

With respect to ASU's, this study found that L2 speakers and L1 speakers do not differ at ASU-boundaries: the likelihood of both silent and filled pauses is not significantly different. However, within ASU's, L2 speakers are more likely to pause – either with a filled or with a silent pause – than L1 speakers are. Whereas Riazantseva (2001) did not find any significant differences with respect to pause distribution for low and high proficient speakers, the results of the current study did show a mediating effect of (gradient) L2 proficiency: the interaction (for silent pauses only) between proficiency and location with respect to ASU showed that as learners were more proficient, they produced fewer silent pauses within ASU's.

The study also investigated whether pauses are more likely to occur before low-frequency words in (semi-)spontaneous speech. Previous research has shown that at least in constrained description tasks (in which participants had to describe a path from picture to picture on a computer screen), L1 speakers are more likely to pause before lower frequency nouns in comparison to higher frequency nouns (Hartsuiker and Notebaert 2010; Kircher et al. 2004). The current study showed that, in more spontaneous speech, both L1 and L2 speakers are more likely to pause before lower frequency nouns than before higher frequency nouns and that L2 proficiency did not modulate this effect. It is likely, however, that the range of which words are considered to be "lower-frequent" and therefore more likely to induce a pause, differs for the L1 and L2 speakers. This option is in fact borne out by the data: it was found that the nouns used by L2 speakers were, overall, higher in frequency than those by L1 speakers. The relation between frequency of nouns and likelihood to pause may be the same (or at least, in this corpus we were unable to prove that the relation was different), but the relation seems to hold for a slightly different range of frequencies: somewhat higher frequency nouns for L2 speakers and the lower frequency nouns for L1 speakers.

The current findings have implications for our understanding of speech production processes in L1 and L2. Because both L1 and L2 speakers pause before ASU's with filled or silent pauses to the same extent, we may infer that at these positions, speakers make their conceptual plan and this causes the speakers to pause, no matter what language they speak. This conclusion is

validated by the finding that between ASU's, L1 and L2 speakers tend to pause for similar durations. Before starting an ASU, a speaker plans his conceptual message and only the beginning of the linguistic message, both in L1 and in L2. It is within ASU's that pausing behaviour differs for L1 and L2 speech. Within ASU's, L2 speakers more often run into trouble while formulating the linguistic message than L1 speakers do and are therefore more likely to pause, for longer durations than L1 speakers do. The pauses within ASU's that L2 speakers use (in excess of those that L1 speakers would use), we may hypothesize, are due to less L2 knowledge and lower L2 skills. Again, this is corroborated by the moderating effect we found for the proficiency of the L2 speakers: the higher the L2 proficiency, the less likely speakers were to pause within ASU's and the shorter the pauses within ASU's tended to be.

This study also found an effect of word frequency on the likelihood to pause: L1 speakers are more likely to pause before lower frequency nouns as compared to higher frequency nouns. This may be explained by postulating that speakers may have trouble while retrieving low frequency referents. The likelihood that speakers run into trouble when retrieving low frequency words is of the same order of magnitude for L1 and L2 speakers, of different levels of proficiency (as no interactions for the effect were found).

It should be noted that the current study is limited in that only intermediate to advanced learners of Dutch participated. It is unclear whether beginner to intermediate learners may show differential effects. Additionally, with the current sample of L1 Turkish and L1 English speaking in their L2 Dutch it is difficult to draw conclusions on differential effects for different language backgrounds. In some cases, we found an effect of language background. Firstly, compared to English speakers, Turkish speakers tended to use more filled pauses within utterances (and before nouns, also more silent pauses). Secondly, Turkish speakers' silent pauses within utterances tended to be shorter. It is beyond the scope of this paper, however, to investigate such potential differences further.

In spite of these limitations, we may conclude that L2 speakers pause more often and for longer durations within utterances than L1 speakers do; at utterance boundaries, the current study did not find any differences. This finding is compatible with the claim that L1 speakers are able to speak 'one clause at a time', but that L2 speakers may need to pause to formulate their message mid-utterance. For the language testing practice, these findings have implications in that (at least for intermediate to advanced learners) pauses between utterances are not informative because they reflect conceptual planning. It is only the pauses that occur within utterances that are indicative of L2 proficiency and are therefore informative for distinguishing on the basis of

aspects of fluency. This suggests that both human judgements and automatic scores in language testing should take pause locations into account. As a matter of fact, human judges will already partly do this even without having been told to do so. For instance, Butcher (1981) showed that pauses at junctures need to be longer in order to be perceived by at least 75 % of the listeners (> 220 ms) than pauses within utterances (> 80 ms).

References

- Baayen, R. Harald, Richard Piepenbrock & Leon Gulikers. 1995. The Celex lexical database (CD-ROM).
- Bates, Douglas, Martin Maechler & Ben Bolker. 2012. lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-39.
- Beglar, David & Alan Hunt. 1999. Revising and validating the 2000 word level and university word level vocabulary tests. *Language Testing* 16(2). 131–62.
- Bortfeld, Heather, Silvia D. Leon, Jonathan E. Bloom, Michael F. Schober & Susan E. Brennan. 2001. Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech* 44(2). 123–47.
- Butcher, Andrew. 1981. Pause and syntactic structure. In Hans W. Dechert, Manfred Raupach (eds.), *Temporal variables in speech*. 85–90. The Hague: Mouton.
- Davies, Alan. 2003. *Native speaker: Myth and reality*. Clevedon: Multilingual Matters.
- De Jong, Nivja H. & Hans Rutger Bosker. 2013. Choosing a threshold for silent pauses to measure second language fluency. In Robert Eklund (ed.), *Proceedings of disfluency in spontaneous speech diss*, 17–20. Stockholm: Royal Institute of Technology (KTH).
- De Jong, Nivja H., Margarita P. Steinel, Arjen F. Florijn, Rob Schoonen & Jan H. Hulstijn. 2012. Facets of speaking proficiency. *Studies in Second Language Acquisition* 34. 5–34.
- De Jong, Nivja H., Rachel Groenhout, Rob Schoonen & Jan H. Hulstijn. 2015. Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics* 36(2). 223–243.
- De Jong, Nivja H., Margarita P. Steinel, Arjen F. Florijn, Rob Schoonen & Jan H. Hulstijn. 2013. Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics* 34. 893–916.
- Derwing, Tracy M., Murray J. Munro, Ron I. Thomson & Marian J. Rossiter. 2009. The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition* 31(04). 533–57.
- Foster, Pauline, Alan Tonkyn & Gillian Wigglesworth. 2000. Measuring spoken language: A unit for all reasons. *Applied Linguistics* 21(3). 354–75.
- Goldman-Eisler, Frieda. 1958. Speech production and the predictability of words in context. *Quarterly Journal of Experimental Psychology* 10. 96–106.
- Hartsuiker, Robert J. & Lies Notebaert. 2010. Lexical access problems lead to disfluencies in speech. *Experimental psychology* 57(3). 169.
- Housen, Alex & Folkert Kuiken. 2009. Complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics* 30(4). 461–73.
- Hox, Joop. 2010. *Multilevel analysis: Techniques and applications*. New York: Routledge.

- Iwashita, Noriko, Annie Brown, Tim McNamara & Sally O'Hagan. 2008. Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics* 29(1). 24–49.
- Kircher, Tilo T., Michael J. Brammer, Willem Levelt, Mathias Bartels & Philip K. McGuire. 2004. Pausing for thought: engagement of left temporal cortex during pauses in speech. *NeuroImage* 21(1). 84–90.
- Maclay, Howard & Charles E. Osgood. 1959. Hesitation phenomena in spontaneous English speech. *Word* 15. 19–44.
- MacWhinney, Brian. 2000. *The CHILDES Project: Tools for analyzing talk*. Mahwah, NJ: Erlbaum.
- Pawley, Andrew & Frances H. Syder. 1983. Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In Jack Richards & Richard Schmidt (eds.), *Language and communication*, 191–226. London: Longman.
- Pinheiro, José C. & Douglas M. Bates. 2000. *Mixed-effects models in S and S-PLUS*. New York: Springer Verlag.
- Riazantseva, Anastasia. 2001. Second language proficiency and pausing. *Studies in Second Language Acquisition* 23. 497–526.
- Riggenbach, Heidi. 1991. Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes* 14(4). 423–41.
- Schmidt, Richard. 1992. Psychological mechanisms underlying second language fluency. *Studies in second language acquisition* 14. 357–.
- Skehan, Peter & Pauline Foster. 2007. Complexity, accuracy, fluency and lexis in task-based performance: A meta-analysis of the Ealing Research. In Siska Van Daele, Alex Housen, Folkert Kuiken, Michel Pierrard & Ineke Vedder (eds.), *Complexity, accuracy, and fluency in second language use, learning, and teaching*. 207–226. Brussels: University of Brussels Press.
- Swerts, Marc. 1998. Filled pauses as markers of discourse structure. *Journal of Pragmatics* 30(4). 485–96.
- Tavakoli, Parvaneh. 2011. Pausing patterns: differences between L2 learners and native speakers. *ELT Journal* 65(1). 71–9.
- Towell, Richard, Roger Hawkins & Nives Bazergui. 1996. The development of fluency in advanced learners of French. *Applied Linguistics* 17(1). 84–119.
- Zareva, Alla, Paula Schwanenflugel & Yordanka Nikolova. 2005. Relationship between lexical competence and language proficiency: Variable sensitivity. *Studies in Second Language Acquisition* 27(4). 567–95.