

EXPLORING THE PRODUCTION AND PERCEPTION OF SECOND LANGUAGE
FLUENCY: UTTERANCE, COGNITIVE, AND PERCEIVED FLUENCY

By

Ji Min Kahng

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Second Language Studies – Doctor of Philosophy

2014

ABSTRACT

EXPLORING THE PRODUCTION AND PERCEPTION OF SECOND LANGUAGE FLUENCY: UTTERANCE, COGNITIVE, AND PERCEIVED FLUENCY

By

Ji Min Kahng

Fluency is one of the most noticeable differences between native and nonnative speech and constitutes an essential component of second language proficiency; however, the concept has not been well understood by researchers. In order to deepen understanding of the multidimensional construct of fluency, the current dissertation investigated the production and perception of second language fluency from all three aspects—utterance, cognitive, and perceived fluency.

Study 1 investigated utterance fluency and cognitive fluency of English speakers and Korean learners of English by comparing temporal measures and stimulated recall responses. The first language (L1) and second language (L2) speech were different in speed, length of run, repairs, and silent pauses. In particular, a striking group difference was found in the frequency of silent pauses within a clause, which is consistent with the claim that pauses within clauses reflect processing difficulties in speech production such as lexical retrieval. Stimulated recall responses showed that lower proficiency learners remembered more issues regarding L2 declarative knowledge on grammar and vocabulary than higher proficiency learners, which was compatible with the declarative/procedural model and studies on automaticity.

Study 2 examined the relationship between utterance fluency and perceived fluency using two experiments. Experiment 1 investigated the relative contributions of frequency, length, and distribution of silent pauses to perceived fluency of L2 speech. Experiment 2 tested causal effects of pause location on perceived fluency of L1 and L2 speech. Findings of both Experiment

1 and Experiment 2 suggest a significant role of pause location in L2 perceived fluency. In Experiment 1, pause distribution demonstrated the strongest correlation with fluency ratings, and in Experiment 2, perceived fluency of L2 speech was influenced by pause location more than that of L1 speech. The findings are consistent with L1 literature on pause phenomena which has shown that silent pauses are one of the acoustic cues to clausal units, and silent pauses between clauses can facilitate speech perception and recall, whereas pauses within clauses can interfere with them in cognitively demanding contexts.

One of the most novel and important findings of the current dissertation is the close relationship between L2 fluency and pauses within clauses. L1 and L2 speech exhibited a striking difference in the frequency of pauses within clauses, which is considered to reflect difficulties in speech production processing. Pauses within clauses also had a crucial impact on perceived fluency of L2 speech.

To Jun

ACKNOWLEDGEMENTS

My deepest gratitude goes to my advisor, Dr. Debra Hardison. I am indebted to her utmost trust in my potential and her constant encouragement throughout my Ph.D. study. Her dedication to students will always be my inspiration.

I would like to thank my committee members. I am immensely grateful to Dr. Susan Gass for both financial support and insightful feedback through various channels, such as practice talks for conference presentations, annual report letters, and defenses. I am deeply thankful to Dr. Patti Spinner for constructive comments on my second qualifying paper and the current dissertation. Her comments allowed me to engage more critically with my research. Dr. Paula Winke provided me with concrete advice, not only on research, but also on life. I truly admire her genuine care for students.

Dr. Aline Godfroid also offered me thought-provoking comments and suggestions. I am grateful for her support and contributions. I appreciate Dr. Karthik Durvasula's inspirational enthusiasm for research and teaching. In addition to the support of my mentors and colleagues, my family and friends have always been there for me through this process, and their unwavering encouragement helped me pull through challenging times. Most importantly, I thank my husband, Jun, for believing in me, inspiring me to pursue my passion with courage, and taking this journey with me.

I greatly appreciate the financial support I received for this dissertation, including funding from the Second Language Studies program, a Dissertation Completion Fellowship from the Graduate School of Michigan State University, and a Dissertation Grant from *Language Learning*. Finally, I thank all the individuals whose participation made this work possible.

TABLE OF CONTENTS

LIST OF TABLES	viii
LIST OF FIGURES	ix
CHAPTER 1: INTRODUCTION AND REVIEW OF THE LITERATURE	1
Introduction	1
Utterance Fluency and Perceived Fluency	2
Speed and Repair Fluency	2
Breakdown Fluency	4
Cognitive Fluency and Utterance Fluency	6
Overview of Research Design	10
CHAPTER 2: STUDY 1	12
Introduction	12
Method	15
Participants	15
Tasks and Procedures	16
Spontaneous Speech	16
Stimulated Recall	17
Utterance Fluency: Quantitative Study	17
Analysis	17
Transcribing and Marking Pauses and Repairs	17
Calculating Temporal Variables	19
Statistical Analysis	22
Results	23
Overall Utterance Fluency	23
Pause Phenomena in Different Locations	24
Correlation Between the Temporal Variables and Speaking Scores	28
Cognitive Fluency: Qualitative Study	31
Analysis	31
Results	32
Content of the Message	33
Vocabulary	34
Grammar	35
Phonology and Pragmatics	36
Other Issues	37
Discussion	38
Utterance Fluency	38
Speed	38
Length of Run	40
Repairs	41
Pause Phenomena	42

Cognitive Fluency	46
General Discussion: Study 1	50
CHAPTER 3: STUDY 2	54
Introduction	54
Experiment 1	55
Method	56
Raters	56
Stimulus Description	56
Procedure	57
Acoustic Analysis of Speech Excerpts	58
Statistical Analysis	59
Results	59
Discussion	62
Experiment 2	64
Method	66
Raters	66
Stimulus Description	67
Procedure	69
Analysis	70
Results	71
Discussion	74
General Discussion: Study 2	77
CHAPTER 4: CONCLUSION	80
APPENDICES	83
Appendix A: Questions for spontaneous speech	84
Appendix B: English language learning background questionnaire in Study 1	85
Appendix C: Instructions for the experiment	86
Appendix D: Rater background questionnaire in Study 2	87
Appendix E: An example of addition of pauses to a speech sample	89
Appendix F: Example waveforms of the speech manipulations	90
REFERENCES	91

LIST OF TABLES

Table 1: Temporal measures used in the study.....	21
Table 2: Overall utterance fluency: Descriptive statistics and group differences	24
Table 3: Mean length of silent and filled pause in different locations (ms)	27
Table 4: Pearson correlations between utterance fluency measures and speaking scores	30
Table 5: Distribution of stimulated recall responses.....	33
Table 6: Summary of statistical analyses.....	39
Table 7: Number of silent and filled pauses per minute in different locations.....	45
Table 8: Descriptive statistics of pause phenomena and fluency ratings of L2 speech.....	60
Table 9: Correlations between the measures of pause phenomena and fluency ratings	60
Table 10: Results of a hierarchical multiple regression.....	61
Table 11: Results of a stepwise multiple regression.....	62
Table 12: A schematic representation of the 3 x 3 Latin Square design. No, B, and W represent the <i>No Pause</i> , <i>Pauses Between Clauses</i> , and <i>Pauses Within Clauses</i> conditions, respectively. ..	69

LIST OF FIGURES

Figure 1: Segalowitz’s (2010) model of the L2 speaker. Adapted from Levelt’s (1999) “blueprint” of the monolingual speaker. The { <i>f</i> } symbols refer to fluency vulnerability points. (Figure used with permission of Taylor and Francis Group LLC Books).....	7
Figure 2: Distribution of speaking and pause time	25
Figure 3: Pause rates for silent and filled pauses in different locations (means and standard errors)	26
Figure 4: Mean and standard error z-scores of fluency ratings of L1 and L2 speech.....	72
Figure 5: Speech in the <i>No Pause</i> condition.....	90
Figure 6: Speech in the <i>Pauses Between Clauses</i> condition.....	90
Figure 7: Speech in the <i>Pauses Within Clauses</i> condition.....	90

CHAPTER 1: INTRODUCTION AND REVIEW OF THE LITERATURE

Introduction

One of the most noticeable differences between speech in first language (L1) and second language (L2) is found in fluency (Gut, 2009; Kormos, 2006). Compared to L1, people are typically not only weaker in L2 knowledge, but they are considerably less fluent using what L2 knowledge they have (Segalowitz, 2010). Although fluency is considered important by L2 learners and teachers (Schmidt, 2000) and constitutes an essential criterion in assessing L2 performance and proficiency (Bosker, Pinger, Quené, Sanders, & De Jong, 2013; Cucchiaroni, Strik, & Boves, 2002; Housen, Kuiken, & Vedder, 2012; Iwashita, Brown, McNamara, & O'Hagan, 2008; Skehan, 1998), the concept is difficult to define and it has not been well understood (Kormos & Dénes, 2004; Schmidt, 2000; Segalowitz, 2010). Lennon (1990) distinguished between fluency in the broad and in the narrow sense. Fluency in the broad sense refers to global speaking proficiency, whereas fluency in the narrow sense relates to how easily and smoothly speech is delivered and it constitutes a component of oral proficiency. The present study concerns the narrow sense of fluency.

Segalowitz (2010) pointed out that even this narrow sense of fluency is a multidimensional construct which reflects the efficiency of using linguistic knowledge and executing the neurological and muscular mechanisms that an L2 speaker has developed during the course of L2 learning, and a distinction should be made among the three notions of fluency—cognitive, utterance, and perceived fluency. Cognitive fluency is defined as “the efficiency of operation of the underlying processes responsible for the production of utterances” (Segalowitz, 2010, p. 165). Utterance fluency is “the features of utterances that reflect the speaker’s cognitive

fluency” (p. 165). Utterance fluency can be objectively measured by temporal variables in speech samples and it has a few different aspects such as *speed fluency*, *breakdown fluency* (pause and hesitation phenomena), and *repair fluency* (Skehan, 2003, 2009; Tavakoli & Skehan, 2005). The third notion of fluency is perceived fluency, “the inferences listeners make about speakers’ cognitive fluency based on their perceptions of their utterance fluency (p. 165).”

Utterance Fluency and Perceived Fluency

Speed and Repair Fluency

To identify reliable oral production features of L2 fluency, previous studies compared speech from fluent and non-fluent speakers (Ejzenberg, 2000; Riazantseva, 2001; Riggenbach, 1991; Tavakoli, 2011), investigated the longitudinal development of fluency (Derwing, Munro, & Thomson, 2007; Freed, 1995, 2000; Lennon, 1990, Mora & Valls-Ferrer, 2012; Towell, Hawkins, & Bazergui, 1996; Wood, 2010), and related utterance fluency to perceived fluency by correlating fluency ratings with temporal variables (Bosker et al., 2013; Cucchiarini, Strik, & Boves, 2000, 2002; Derwing, Rossiter, Munro, & Thomson, 2004; Fulcher, 1996; Kormos & Dénes, 2004; Rossiter, 2009).

The main findings were that *Speech rate* (i.e., the number of syllables per minute, including pause time) and *Mean length of run* (i.e., the mean number of syllables between two silent pauses) were consistently strongly associated with L2 oral fluency development and perceived fluency (e.g., Kormos & Dénes, 2004; Lennon, 1990; O’Brien, Segalowitz, Freed, & Collentine, 2007; Segalowitz & Freed, 2004; Towell et al., 1996), whereas *Articulation rate* (i.e., the number of syllables per minute, excluding pause time) and repair measures often did not. For instance, Kormos and Dénes (2004) compared temporal features of speech produced by

intermediate and advanced learners of English and found that they were significantly different in *Speech rate* and *Mean length of run* but not in *Articulation rate* or *Number of disfluencies* (repetitions, restarts and repairs) *per minute*. Correlation analysis between the temporal measures and fluency scores rated by native English speakers showed that *Speech rate* and *Mean length of run* very strongly correlated with fluency scores, while *Articulation rate* and *Number of disfluencies per minute* did not. Cucchiarini et al. (2002) also found that fluency ratings of speech samples produced by beginning and intermediate learners of Dutch were moderately to strongly correlated with *Speech rate* and *Mean length of run*; on the other hand, fluency ratings were not correlated with *Articulation rate* or *Number of disfluencies per minute*.

However, in Towell et al. (1996) advanced learners of French improved *Articulation rate* after a year abroad even if the improvement in *Articulation rate* was smaller than that in *Speech rate*. Ginther, Dimova and Yang (2010) investigated relationships between oral English proficiency and temporal measures of fluency and found that *Articulation rate* strongly correlated with speaking scores, although less strongly than *Speech rate* did. Considering that L2 fluency can be affected by L1 fluency and speaking style, De Jong, Groenhout, Schoonen and Hulstijn (2013) examined whether correcting measures of L2 fluency for L1 performance can better predict L2 proficiency. They found that L2 *Mean syllable duration* (i.e., the inversion of *Articulation rate*) was able to explain 30% of variance of L2 proficiency and correcting the measure by partialing out L1 behavior increased the explained variance to 41%. Regarding repair measures, Bosker et al. (2013) recently investigated relative contributions of speed, pauses, and repairs to perceived fluency by examining perceptual sensitivity to the three fluency aspects, and found repairs did contribute a small but significant amount to perceived fluency. Therefore,

whether *Articulation rate* and repair measures are indicators of L2 fluency or not is not yet conclusive.

Breakdown Fluency

Previous findings on pause phenomena show an even more complicated picture. In Ginther et al. (2010) and Bosker et al. (2013), both pause frequency and pause length were negatively correlated with proficiency scores and fluency ratings, respectively. On the other hand, in Kormos and Dénes (2004) fluency ratings did not correlate with pause frequency but did correlate with pause length. Furthermore, in Cucchiari et al. (2002), the opposite pattern was found, in other words, fluency ratings correlated with pause frequency but not with pause length.

Pause distribution has been investigated by even fewer studies, in which fluent speech tended to have pauses at grammatical junctures (Lennon, 1990; Towell et al., 1996), whereas non-fluent L2 speech often had pauses within clauses (Davies, 2003; Deschamps, 1980; Freed, 1995; Riggenbach, 1991; Tavakoli, 2011). It has been argued that in fluent speech, language is encoded a clause at a time (Pawley & Syder, 2000) and pausing within clauses seems to reflect difficulties in planning or encoding speech (Cenoz, 1998; Lennon, 1984; Wood, 2010). However, the results are not yet conclusive because as Kormos (2006) points out, many earlier studies suffer from very small sample sizes (often with a few to several participants) and did not use computer technology to obtain more precise temporal measures in milliseconds or statistical analyses, suggesting difficulty in generalizing the results. Moreover, in Riazantseva (2001) which had a higher number of participants (20 L1 speakers; 30 L2 speakers in total, 15 per group), no difference was found in the number of within-constituent pauses between L1 and L2 speakers.

L1 literature has a longer history on pause phenomena (e.g., pausology—a specialized field in psycholinguistics, the study of temporal variables in speech pioneered by Goldman-Eisler in the 1950s) and can provide insights on L2 research (Griffiths, 1991). Schnadt (2009) points out that one of the major issues for the study of silent pauses has been distinguishing a “hesitant” pause from a pause based on a speaker’s natural prosody. Hesitant pauses (or performance-based pauses, Ferreira, 1993, 2007) are related to delays in planning and production processes, whereas prosodic pauses (Ferreira, 1993, 2007) separate utterances into intonational phrases (i.e., a speech segment which occurs with a single prosodic contour), and thus are part of the rhythmic structure of speech. Indeed, in L1 speech, most pauses tend to occur at clause boundaries (Boomer, 1965; Hawkins, 1971; Holmes, 1988; MacGregor, 2008). Prosodic pauses typically occur at intonational phrase or clause boundaries; however, performance-based pauses can occur at any point where a speaker needs to plan upcoming speech or encounters difficulty.

L1 research on pause phenomena has also shown its important role in speech perception and comprehension. Silent pauses at grammatical boundaries have been claimed to help listener comprehension as they enable them to understand and keep pace with the utterance by indicating the boundaries of speech to be analyzed, and providing cognitive processing time (Arons, 1993; Griffiths, 1991; Reich, 1980, Sugito, 1990). Pauses at grammatical junctures are important for comprehension and eliminating them can interfere with comprehension (Lass & Leeper, 1977). However, as Arons (1993) maintains, only pauses between clauses or structural pauses (i.e., pauses between items of information in lists of meaningful trigrams such as IBM [pause] KGB [pause] PHD) are useful; pauses within clauses or nonstructural pauses (e.g., DIB [pause] MKG [pause] BPH) can interfere with speech perception processing (Bower & Springston, 1970; Griffiths, 1991; Reich, 1980; Sugito, 1990). Silent pauses are one of the acoustic cues to clausal

units along with pitch and vowel duration (Seidl & Cristià, 2008) and in language development of infants, by 6 months of age, infants show a preference for sentences containing pauses between clauses to sentences containing pauses within clauses (Hollich & Houston, 2007). In Reich (1980), words were categorized faster and propositions were recalled more accurately in sentences containing pauses between clauses than in sentences containing pauses within clauses. It has also been reported that silent pauses have beneficial effects on listeners under conditions of cognitive complexity in auditory speech processing and they did not demonstrate apparent beneficial effects when the speech or tasks were easy enough (Aaronson, 1968; Reich, 1980).

Cognitive Fluency and Utterance Fluency

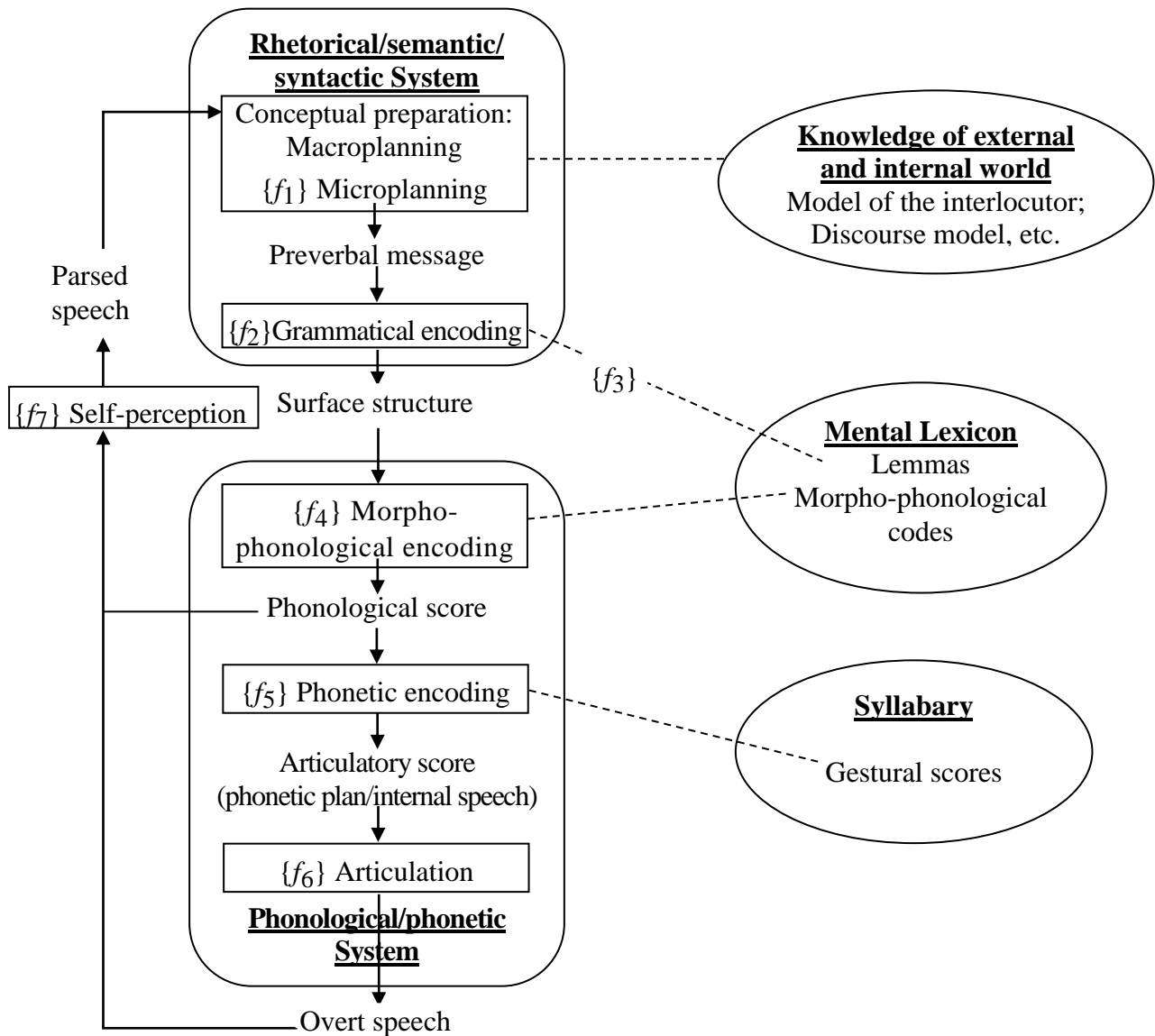
In understanding the underlying mechanisms responsible for L1 and L2 oral fluency, the differences are often explained by the degree of automaticity. Automaticity refers to the absence of attentional control in executing a cognitive activity (Kahneman, 1973) and has several characteristics, such as rapidity, effortlessness, unconscious and ballistic nature (Segalowitz & Hulstijn, 2005). Kormos (2006) points out that whereas L1 speech production requires attention only to speech planning and monitoring, in L2 speech, syntactic and phonological encoding may not be fully automatized, slowing speech down.

For systematic understanding of L2 cognitive fluency, as shown in Figure 1, Segalowitz (2010) adopted Levelt's (1999) L1 speech production model and identified possible fluency vulnerability points (i.e., critical points where underlying processing difficulties could result in L2 speech disfluencies, {f} symbols) by incorporating De Bot's (1992) proposals on bilingual speakers (see Segalowitz, 2010, pp. 7-17 for details).

Figure 1: Segalowitz’s (2010) model of the L2 speaker. Adapted from Levelt’s (1999)

“blueprint” of the monolingual speaker. The {f} symbols refer to fluency vulnerability points.

(Figure used with permission of Taylor and Francis Group LLC Books)



Seven fluency vulnerability points were proposed, microplanning ({f1}), grammatical encoding ({f2}), lemma retrieval ({f3}), morpho-phonological encoding ({f4}), phonetic

encoding (f_5), articulation (f_6), and self-perception (f_7). During macroplanning (i.e., planning what to say next) no L2-specific fluency issues are expected, as world knowledge used in this stage is not assumed to be organized in language specific terms (De Bot, 1992; Levelt, 1989, 1999).

In microplanning, language-specific information (e.g., argument structure, mood, and tense-aspect) is included in a preverbal message and L2 speakers sometimes strategically formulate a preverbal message to avoid L2 difficulties (De Bot, 1992). Taking one's limitation into consideration may slow down the process of formulating the preverbal message (f_1). Lexical retrieval is one of the most salient problems with L2 speakers (De Bot, 1992) and it seems difficult for them to retrieve (f_3) and utilize the appropriate linguistic resources to create a grammatical surface structure (f_2). Whereas the process is highly automatized in L1 (Levelt, 1989), in morpho-phonological encoding (i.e., specifying morphological, segmental, and suprasegmental structure of the word), L2 learners may not have automatic access to syllable programs (f_4). L1 and L2 are likely to utilize different repertoires of articulatory gestures and L2 fluency can be compromised when speakers cannot automatically access appropriate gestural scores (f_5), or execute the score (f_6).

Additionally, Kormos (2006) included an L2 declarative knowledge store in the bilingual speech production model based on the declarative/procedural model (Ullman, 2001, 2004, 2005, 2013). According to the model, in L1, lexicon and grammar are learned, represented, and processed in two memory systems—declarative memory and procedural memory. Lexical aspects depend on declarative memory, which is implicated in the learning/use of facts and

declarative memory can be explicitly recollected (Ullman, 2001, 2004, 2005, 2013). On the other hand, grammatical aspects depend upon procedural memory, which is implicated in the learning/use of motor/cognitive skills. Procedural memory is often referred to as an “implicit memory system” as the knowledge itself or procedures of the learning of the knowledge is not accessible to conscious memory (Ullman, 2004). Learning in the procedural system results in rapid and automatic processing of skills and knowledge (Ullman, 2013). Ullman claims that in late L2 learning after puberty, however, grammatical computation as well as the mental lexicon largely relies on declarative memory, and L2 experience (practice) and age of exposure to L2 affect the relative reliance on declarative versus procedural memory. The claim has been supported by neuroimaging studies (e.g., Dehaene et al., 1997; Perani et al., 1998), which found a greater activation in the brain regions responsible for declarative memory in L2 (the hippocampus and medial temporal lobe structures) for the processing of forms which are mainly computed by the procedural memory in L1. In addition, the activation pattern from early bilinguals was similar to L1 speakers, whereas low-proficiency L2 speakers heavily relied on the declarative memory system (Perani et al., 1998). Opitz and Friederici (2003) also found that when adults were learning an artificial language, during syntactic processing the hippocampus and the temporal lobe were activated; the areas are involved in declarative memory. The findings suggest that declarative knowledge of grammar is stored in a different area from where procedural knowledge is stored, and as learners are exposed to L2 at an earlier age and as they practice more, they become more dependent upon the procedural memory system (Ullman, 2001, 2004). Therefore, in L2, many of the syntactic, lexical and phonological rules which are not automatized are considered to be stored as declarative knowledge (Kormos, 2006; Ullman, 2001).

Until now, only a few studies have investigated cognitive fluency in relation to utterance fluency by measuring subprocesses of speech production. Segalowitz and Freed (2004) measured cognitive fluency using a semantic classification task and an attention control test, and related the results to gains in utterance fluency. They found correlations between mean length of run and lexical access speed and efficiency. De Jong, Steinel, Florijn, Schoonen, and Hulstijn (2013) also aimed to identify utterance fluency measures indicative of cognitive fluency. They measured linguistic knowledge (e.g., vocabulary, grammar, pronunciation) and processing skills (e.g., speed of morphosyntactic processing, lexical selection, articulation). Results showed that linguistic knowledge and skills were most strongly related to average syllable duration, explaining 50% of the variance, whereas they had the weakest correlation with pause duration.

Overview of Research Design

Taken together, previous studies on L2 fluency mainly investigated utterance fluency and perceived fluency in second language. Only a few studies looked at cognitive fluency in relation to oral fluency. Results of utterance and perceived fluency seem to suggest that speed and pause phenomena be related to perception of fluency; however, when closely examined, results are mixed depending on how each aspect of fluency is measured (e.g., speech rate vs. articulation rate) and the relative contribution of frequency, duration, and distribution of pauses on fluency perception have not been fully investigated. Furthermore, there are gaps in the literature on how L1 and L2 speakers' fluency differs (e.g., use of filled pauses, distribution of pauses).

Mixed results in the previous studies may be due to methodological issues such as variability in speech elicitation tasks, in cut-off points for pauses, and small sample sizes. But more importantly, a lack of comprehensive theoretical framework to understand fluency

(Segalowitz, 2010) may have been a barrier to systematic investigation of L2 fluency. Therefore, within Segalowitz's (2010) fluency framework, the current dissertation aims to investigate 1) utterance fluency and cognitive fluency by examining in what respects and why native speakers' and L2 speakers' oral fluency are different in Study 1 and 2) perceived fluency by testing effects of frequency, duration, and location of pauses on the perception of fluency in Study 2.

CHAPTER 2: STUDY 1

Introduction

Study 1 investigated (1) utterance fluency, by comparing speech samples from L1 English speakers and L1 Korean L2 English speakers to show in what respects they are different, and (2) cognitive fluency, by examining stimulated recall responses from L2 speakers to gain an insight regarding the underlying processes and problems related to L2 disfluencies.

Following Skehan's (2003) taxonomy, utterance fluency was investigated in terms of speed, repairs and breakdown fluency. As mixed results had been found in the literature on breakdown fluency, pause phenomena were examined rigorously by investigating frequency, duration, and distribution of both silent and filled pauses. Especially, as pause distribution has rarely been the main focus of L2 fluency research, frequency and duration of silent and filled pauses in different locations (e.g., within clauses, at clause boundaries) were analyzed in depth. In addition, length of run was also included in the analysis of utterance fluency with consideration for its strong association with L2 oral fluency development (e.g., Ginther et al., 2010; Kormos & Dénes, 2004; Towell et al., 1996) and its conceptual connection with automatic speech production processing. As discussed earlier, length of run has been consistently found to be correlated with L2 oral fluency development and perceived fluency in previous studies. Towell et al. (1996) claimed that increase in length of run reflects proceduralization of declarative knowledge based on Anderson's (1983) ACT* (Adaptive Control of Thought) model of skill acquisition. Following Towell et al. (1996) and Towell and Dewaele (2005), Skehan (2009) suggested that length of run can be measured as an indicator of the degree of

automatization in speech performance¹. Length of run seems to be related to automaticity and automatization as automatic speech production processes are likely to result in long fluent runs. Furthermore, length of run also seems to be closely related to the use of prefabricated language units and formulaic language, which have been claimed to facilitate L2 oral fluency (Boers, Eyckman, Kappel et al., 2006; Bybee, 2002; Kuiper, 1996; Skehan, 1998).

The study used stimulated recall to investigate cognitive processes of L2 speakers regarding disfluencies. As mentioned above, only a couple of studies (De Jong, Steinel, et al., 2013; Segalowitz & Freed, 2004) have investigated cognitive fluency in relation to utterance fluency and they tried to measure cognitive processes involved in cognitive fluency, for instance, by measuring attention control and speed of morphosyntactic processing. None of them utilized stimulated recall, whereas it was used in a few studies on problem-solving mechanisms and self-monitoring (e.g., Dörnyei & Kormos, 1998; Kormos, 2000a, 2000b). Stimulated recall is different from other cognitive measures used in the studies on cognitive fluency in that it is a more global and indirect measure of cognitive processes. Stimulated recall can reflect cognitive events and reveal the information attended to during task performance (Gass & Mackey, 2000). Stimulated recall has a limitation in capturing subconscious cognitive processes; however, it has the potential to extend our understanding about a complex phenomenon of L2 fluency by tapping

¹ Given that automaticity has been operationalized in various ways in the literature based on its complex characteristics including rapidity, effortlessness, unconscious and ballistic nature (Segalowitz & Hustijn, 2005), it is unlikely that automatization can be measured by a single measure such as length of run. The direct connection between automatization and length of run is also still not clear. For instance, it is unclear whether long fluent run reflects automaticity in lemma retrieval, grammatical encoding, morpho-phonological encoding, phonetic encoding, articulation, or combination of some of these stages. Moreover, it is unlikely that other measures of utterance fluency such as *Mean syllable duration* does not relate to automatization in speech production. Therefore, the measure of length of run is included in the analysis as its own category (see Table 1) and not under the category of automatization (c.f., Koponen & Riggensbach, 2000; Skehan, 2003).

issues that L2 speakers attend to during speaking, which cannot be easily addressed by other methods. Following are the three research questions in Study 1, addressing utterance fluency and cognitive fluency.

1. Are there differences in utterance fluency (i.e., speed, length of run, repairs, and frequency, length, and distribution of pauses) between L1 and L2 speakers?
2. Which of the utterance fluency measures (i.e., speed, length of run, repairs, and frequency, length, and distribution of pauses) are correlated with L2 oral proficiency?
3. Are there differences in cognitive fluency reflected in the stimulated recall responses between lower and higher proficiency learners?

Although Study 1 is primarily exploratory in nature, some predictions are possible to make based on previous studies on fluency. In utterance fluency, if pauses within clauses reflect processing difficulties in speech production as proposed by Lennon (1984), Pawley and Syder (2000) and Wood (2010), L2 learners are expected to pause within clauses more often than L1 speakers. In addition, as L2 proficiency increases, frequency of pauses within clauses is expected to decrease.

Moreover, according to the declarative/procedural model (Ullman, 2001, 2004, 2005, 2013) and previous studies on automaticity (DeKeyser, 2001, 2007; Segalowitz, 2000, 2003), L2 learners tend to rely on declarative memory/knowledge and become more dependent upon procedural memory/knowledge as they practice and gain more experience in L2. Therefore, lower proficiency learners are expected to use L2 declarative knowledge more than higher proficiency learners. Considering that only declarative knowledge can be explicitly recollected

and procedural knowledge cannot (Ullman, 2001, 2004, 2013), lower proficiency learners are expected to remember more about their thoughts at the time of speaking in stimulated recall than higher proficiency learners. Furthermore, in terms of the content of the stimulated recall responses, if higher proficiency learners' production processes are more automatized than lower proficiency learners,' higher proficiency learners are predicted to report on macroplanning and monitoring that required their attention as in L1 speech (Kormos, 2006), whereas lower proficiency learners are expected to report on more varied issues including syntactic and morpho-phonological encoding that are not fully automatized and controlled in the declarative memory system.

Method

Participants

Thirty-one Korean learners of English (10 males; 21 females) and 15 English native speakers (4 males; 11 females) participated in the study. The mean age of the Korean speakers was 31 ($SD = 6.3$) and they started to learn English around the age of 12 ($SD = 0.1$). None of them had any other languages spoken at home as a child other than Korean, or learned English before the age of 8. Their length of residence ranged from 1 month to 7.5 years ($M = 1.9$, $SD = 1.7$).

Korean participants' English oral proficiency was measured by evaluating their speech samples holistically using a rubric of the SPEAK test (Indiana University, n.d.), which had criteria such as clarity and effectiveness of communication, effective use of linguistic features, severity of linguistic errors, organization, and appropriateness (scale range: 20 – 60); the rubric did not entail specific aspects of fluency. Two raters evaluated speech samples and their

interrater reliability measured by Pearson correlation coefficient was $r = .910$ ($p < .001$). The average score was used for each participant ($M = 47$, $SD = 7.5$, $Min. = 34$, $Max. = 60$). The mean age of English speakers was 29 ($SD = 5.7$) and they were undergraduate or graduate students at a university in the United States.

Tasks and Procedures

All 46 participants completed a spontaneous speech task (Appendix A) and a paper-based survey questionnaire on their L2 learning background (Appendix B), and 17 Korean learners of English within this group voluntarily participated in a stimulated recall session. The study was conducted individually in a quiet room with a Korean-English bilingual researcher present. The tasks were presented in *PowerPoint* on a laptop. The participants' speech was recorded through a low-noise headset using the digital audio editing software *GoldWave* and the recordings were saved as 22KHz (16-bit resolution; 1-channel).

Spontaneous Speech

Two questions were used to elicit spontaneous speech, in which the participants spoke freely on a given topic. The questions were on daily life so that all the participants were familiar with the topic and were able to talk naturally without much difficulty. The first question was about their major field of study, what it was about, whether they liked it or not and why. The second one was about their free-time activities (Appendix A).

When each question appeared on the screen, the participants were able to start answering the question whenever they were ready. However, none of them spent more than 10 seconds before they answered. They were asked to respond to each question for one minute but they were not interrupted in the middle of their speech after the requested one minute was over. A stop

watch was placed next to the laptop computer so that the participants were able to check the time. When they finished answering each question, they clicked and moved on to the next question at their own pace.

Stimulated Recall

The 17 Korean speakers (7 males; 10 females) voluntarily participated in the stimulated recall session. The participants were asked to describe what they were thinking while pausing or hesitating during their speech. Stimulated recall was conducted in their L1, Korean. The session was conducted following Gass and Mackey's (2007) recommendations for stimulated recall research. The audio-recorded spontaneous speech was played for each learner immediately after the spontaneous speech task was over in order to utilize recent memory and reduce recall interference. During the session the participants were allowed to pause the audio file whenever they wanted to describe their thoughts at the time of speech production to make their recalls less susceptible to researcher interference. The researcher was also able to pause the audio file after silent or filled pauses in the speech recording to ask participants to recall their thoughts so as not to let the session become completely unstructured and lose useful data.

Utterance Fluency: Quantitative Study

Analysis

Transcribing and Marking Pauses and Repairs

All speech recordings were transcribed in detail including information regarding pauses and repairs. The length of silent and filled pauses was measured in milliseconds (ms) by listening to each speech sample and examining the waveform and spectrogram using Praat (Boersma &

Weenink, 2012) and the duration was added to the transcript. In previous studies the lower bound of silent pauses varied considerably (100 ms, Kang, Rubin, & Pickering, 2010; Riazantseva, 2001; Trofimovich & Baker, 2006; 200 ms, Cucchiariini et al., 2002; 250 ms, De Jong, Groenhout, et al., 2013; De Jong, Steinel, et al., 2013; Ginther et al., 2010; Raupach, 1987; 280 ms, Towell et al., 1996; 300 ms, Wood, 2010; 400 ms, Derwing et al., 2004; Freed, Segalowitz, & Dewey, 2004; O'Brien et al., 2007; Tavakoli, 2011).

In the present study any silence² equal to or longer than 250 ms was identified as a silent pause and included in the analysis following De Jong, Groenhout, et al. (2013) and Goldman-Eisler (1968). Further support for 250 ms over 400 ms, which is another popular choice in L2 fluency studies, came from recent studies by De Jong and Bosker (2013) and Kahng (2012). In De Jong and Bosker (2013), a lower cut-off point for silent pauses of 250-300 ms led to the highest correlation between the number of silent pauses and L2 proficiency scores. Kahng (2012) compared the results of the analysis based on the two cut-off points for silent pauses and found that 400 ms missed 12% of the pauses identified by 250 ms. More importantly, 77% of the pauses which 400 ms missed were pauses within clauses. As pause distribution is one of the main foci of the present study, 250 ms was selected so as not to lose potentially important information. Filled pauses were defined as nonlexical fillers such as *um* and *uh* (Freed, 1995; Kang et al., 2010; Riggenbach, 1991).

Pause distribution was operationalized by categorizing pauses into pauses within clauses, at clause boundaries, or at AS-unit boundaries. A clause was required to consist minimally of a finite or non-finite verb with at least one other clause element such as a subject, object, or complement (see Foster, Tonkyn, & Wigglesworth, 2000, pp. 365-368). An AS-unit (ASU) is a

² Any silent portions preceding or following filled pauses were also counted as silent pauses as long as they were equal to or longer than 250 ms.

single speaker's utterance which consists of either an independent clause, or sub-clausal unit, with any subordinate clause (Foster et al., 2000). The ASU was devised based on the T-unit (i.e., "a main clause plus any other clauses which are dependent upon it" (Foster et al., 2000, p. 360) but elaborated to deal with the features of spoken data. For instance, unlike the T-unit, the ASU includes independent sub-clausal units and minor utterances (e.g., Oh poor woman, Thank you very much, Yes) which are common in speech. Along with clauses, ASUs which are larger than single clauses were used because speakers may plan multi-clause units (Beattie, 1980) and being able to plan multiple clauses seems to be related to L2 proficiency (Foster et al., 2000). In the transcript, an ASU boundary was marked by a double slash ...//... a clause boundary was marked by a double colon :: and repairs such as repetitions and self-corrections were put inside brackets {...}. Silent pauses were marked in parentheses for duration (in milliseconds) and the duration of filled pauses was marked next to each filled pause without parentheses. Following is an example of a transcript with information about clause and ASU boundaries, repairs, and the duration of silent and filled pauses.

When I was in high school :: I joined {this club} drama club (425) //
and I performed in ah410 several plays //
and I believe :: I have some talent in (484) acting //

Calculating Temporal Variables

Table 1 lists the temporal measures and their operational definitions. The choice of measures was made so that the measures clearly represent each aspect of fluency (i.e., speed, length of run, repairs, and pause phenomena) and they are not mathematically dependent or

strongly interrelated. For speed fluency, *Mean syllable duration* was computed, which is the inverse of the *Articulation rate* measure. As De Jong, Steinel, et al. (2013) pointed out, *Mean syllable duration* is a pure measure of speed in that it excludes pause time unlike the traditional measure of *Speech rate* which includes pause time. For length of run, *Mean syllables per run* was used. For repair fluency, *Number of corrections per minute* and *Number of repetitions per minute* were calculated.

To examine pause phenomena (i.e., *breakdown fluency*), frequency of pauses was measured by *Number of pauses per minute* and duration of pauses by *Mean length of pauses*. To measure distribution of pauses, this study devised *Pause rate* in different locations to capture a more accurate picture of pause distribution than a measure used in a previous study such as *Number of pauses per minute within a clause or at a clause boundary* (Tavakoli, 2011). *Pause rate* takes into account the fact that speech samples do not have the same number of clauses or ASUs across speakers, and there is always an equal or a greater number of clause boundaries than ASU boundaries by definition. For instance, when there are 10 ASU boundaries and 16 clause boundaries in a one-minute speech sample and the speaker paused at all 10 ASU boundaries and all 16 clause boundaries, *Number of pauses per minute* can incorrectly suggest that the speaker paused more often at clause boundaries (16) than at ASU boundaries (10), although the speaker paused at every clause and at every ASU boundary. *Pause rate* computes how often a speaker pauses within each clause, at each clause boundary, or at each ASU boundary. Therefore, in the above example *Pause rate* would be 1 for both clause ($16/16 = 1$) and ASU boundaries ($10/10 = 1$). Moreover, as *Pause rate* in different locations (i.e., within a clause, at a clause boundary, and at an ASU boundary) captures how likely a speaker pauses in

each location³, in calculating *Pause rate*, when more than one pause occurred in one location, they were counted as one. For instance, in “I like chocolate :: ah260 (400) um100 because it is sweet,” even if the speaker used three individual pauses at the clause boundary, they were counted as one because all of them occurred in the same location, between “chocolate” and “because.”

Table 1: Temporal measures used in the study

Measures	Submeasures	Definition and formula
Speed	<i>Mean syllable duration</i> (ms)	Speech time excluding pause time / total number of syllables
Repair	<i>Number of corrections per minute</i>	Total number of corrections / spoken time
	<i>Number of repetitions per minute</i>	Total number of repetitions / spoken time
Length of run	<i>Mean syllables per run</i>	Average number of syllables produced between two silent pauses
Pause	<i>Number of pauses per minute</i>	Total number of pauses / spoken time
	<i>Mean length of pauses</i> (ms)	Total length of pause time / number of pauses
	<i>Pause rate within a clause</i>	Total number of pauses within clauses / number of clauses
	<i>Pause rate at a clause boundary</i>	Total number of pauses at clause boundaries / number of clause boundaries
	<i>Pause rate at an ASU boundary</i>	Total number of pauses at ASU boundaries / number of ASU boundaries
	<i>Ratio of filled to silent pause</i>	Total length of filled pause time / total length of silent pause time
	<i>Pause concurrence rate</i>	Number of pauses occurring concurrently / total number of pauses

Note. Spoken time = duration of speech fragment excluding silences of ≥ 250 ms. For the pause measures, silent and filled pauses were measured separately.

In addition, two more measures—*Ratio of filled to silent pause* and *Pause concurrence rate*—were devised in the current study to explore relationships between silent and filled pauses,

³ By definition, it is not possible to pause more than once at a clause boundary or at an ASU boundary, whereas it is possible within a clause. For instance, in “I um200 (350) don’t (300) know,” the speaker paused two times within a clause.

which has rarely been addressed in L2 fluency studies. *Ratio of filled to silent pause* was the ratio of filled to silent pause time to investigate whether there is a difference in the ratio between L1 and L2 speakers. *Pause concurrence rate* was devised based on the observation that filled pauses are often preceded or followed by silent pauses (Beattie, 1977). When a filled pause is used, it usually interrupts the silence and breaks the silence into smaller pieces, suggesting a possibility that L1 speakers can use filled pauses strategically, for instance, to keep the floor (Beattie, 1977; Taboada, 2006) and to make their speech sound less disfluent. *Pause concurrence rate* computes the probability for pauses to occur right next to each other by dividing number of pauses occurring concurrently by total number of pauses.

Number of words and syllables were counted using online software, *Syllable Counter* (SyllableCount.com, n.d.), which is developed based on an English syllable dictionary. Once transcripts of speech recordings were entered into the window of the software, it produced a result table containing the number of words and syllables for the given text. Result tables also showed the words that the software did not recognize (e.g., ah, um, TESOL) and those words and syllables were counted manually by the researcher.

Statistical Analysis

For statistical analysis, multivariate analyses of variance (MANOVAs) were run using SPSS Statistics 17.0 (SPSS Inc., 2008). In reporting the results of MANOVAs, Pillai's trace was used as it is robust to small and unequal sample sizes (Hair, Black, Babin, & Anderson, 2009). The variables which violated the assumptions of parametric tests were transformed using square root transformations (Larson-Hall, 2010). All the transformed data improved in terms of

normality and homogeneity of variances after the transformation⁴. Three MANOVAs were run in order to investigate group differences in terms of 1) overall temporal features, 2) pause rate in different locations, and 3) pause length in different locations. As three separate MANOVAs were run, the alpha level was set at 0.016 after Bonferroni adjustment (0.05/3) for the omnibus tests to control for Type I error. For the follow-up analysis of variance (ANOVA) tests, the alpha level was also adjusted using Bonferroni correction by dividing 0.05 by the number of dependent variables included in each MANOVA test.

Results

To summarize the mean utterances produced by the L1 English speakers and L1 Korean L2 English speakers, the L1 speakers produced 203 words ($SD = 82$), 287 syllables ($SD = 112$), 31 clauses ($SD = 13$), 18 ASUs ($SD = 7.0$) and the L2 speakers produced 168 words ($SD = 62$), 237 syllables ($SD = 86$), 24 clauses ($SD = 9.1$), and 16 ASUs ($SD = 4.7$) in the two speech samples. In the following, the descriptive statistics and group differences in the measures of overall utterance fluency, pause phenomena in different locations, and the results of correlation analysis among the variables are reported in turn.

Overall Utterance Fluency

Table 2 demonstrates descriptive statistics and group differences in the measures of overall utterance fluency. Results of a one-way MANOVA showed that using Pillai's trace, the two groups were significantly different regarding overall utterance fluency, $V = .71$, $F(8, 37) =$

⁴ The transformed variables were *Number of corrections per minute*, *Number of repetitions per minute*, *Number of filled pauses per minute*, *Mean length of silent pauses at a clause boundary*, and *Mean length of silent pauses at an ASU boundary*. As they were moderately positively skewed, square root transformations were used (Larson-Hall, 2010).

11.05, $p < .001$. Separate univariate ANOVAs on the 8 dependent variables revealed significant group differences in *Mean syllable duration*, *Mean syllables per run*, *Number of corrections per minute*, *Number of silent pauses per minute* with large effect sizes (.01 = small, .06 = medium, .14 = large; Cohen, 1988), and approaching significance in *Mean length of silent pauses*, *Number of repetitions per minute*, and *Number of filled pauses per minute*. Therefore, the L1 speakers spoke faster, produced more syllables per run, and used fewer corrections and silent pauses per minute compared to the L2 speakers. However, the two groups were not statistically different in terms of *Mean length of filled pauses*. It is also interesting to note that the L1 speakers used more filled pauses per minute than the L2 speakers, although the difference did not reach significance at the .006 level.

Table 2: Overall utterance fluency: Descriptive statistics and group differences

	L1 Speakers ($N = 15$)		L2 Speakers ($N = 31$)		F	df	p	η_p^2
	M	SD	M	SD				
Mean syllable duration (ms)	249	37	321	45	28.60	1	< .001*	.40
Mean syllables per run	14.3	5.44	6.25	2.03	53.58	1	< .001*	.55
Number of corrections per minute _t	0.67	1.17	2.21	1.57	16.10	1	< .001*	.27
Number of repetitions per minute _t	0.82	0.90	2.06	1.78	6.42	1	.015	.13
Number of silent pauses per minute	15.1	2.85	21.8	4.24	30.43	1	< .001*	.41
Number of filled pauses per minute _t	9.37	3.21	6.51	4.39	6.26	1	.016	.12
Mean length of silent pauses (ms)	685	170	893	269	7.50	1	.009	.15
Mean length of filled pauses (ms)	499	95	563	127	3.02	1	.089	.06

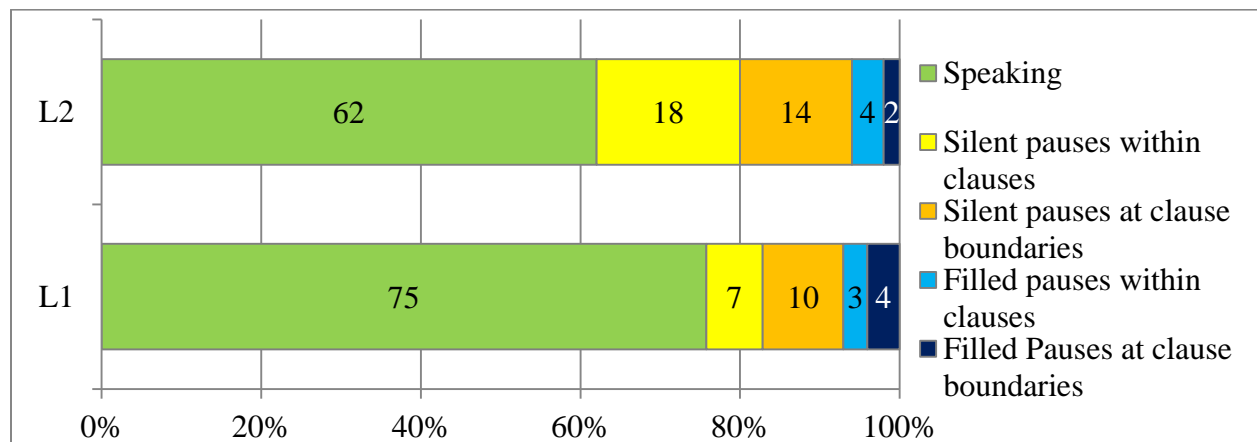
Note. Silent pause ≥ 250 ms. The subscript t next to a variable indicates that the variable was square-root-transformed for inferential statistics. An asterisk indicates significant difference at the 0.006 level after Bonferroni correction (0.05/8).

Pause Phenomena in Different Locations

Figure 2 illustrates the overall distribution of speaking and pause time. As pauses at ASU boundaries are also at clause boundaries, they are included in pauses at clause boundaries and are

not shown separately in the figure. It shows that the L1 speakers used 25% of their response time on pausing, whereas the L2 speakers used 38% of their time on pausing. The striking group difference was found in silent pauses. Silent pause time in L2 speech was almost double compared to that in L1 speech (32% vs. 17%); especially within clauses, the L2 learners spent over twice the amount of time in silent pauses than the L1 speakers did (18% vs. 7%). The figure also illustrates that filled pauses were used much less often than silent pauses by both groups. However, the gap between the two types of pause seems greater for the L2 learners than the L1 speakers. The means and standard deviations of *Ratio of filled to silent pause time* for the L1 speakers was 0.51 (0.32), whereas 0.22 (0.18) for the L2 learners. Although relatively large standard deviations seem to reflect some individual differences in the use of filled pauses, the L2 learners still used silent pauses much more often than filled pauses compared to the L1 speakers.

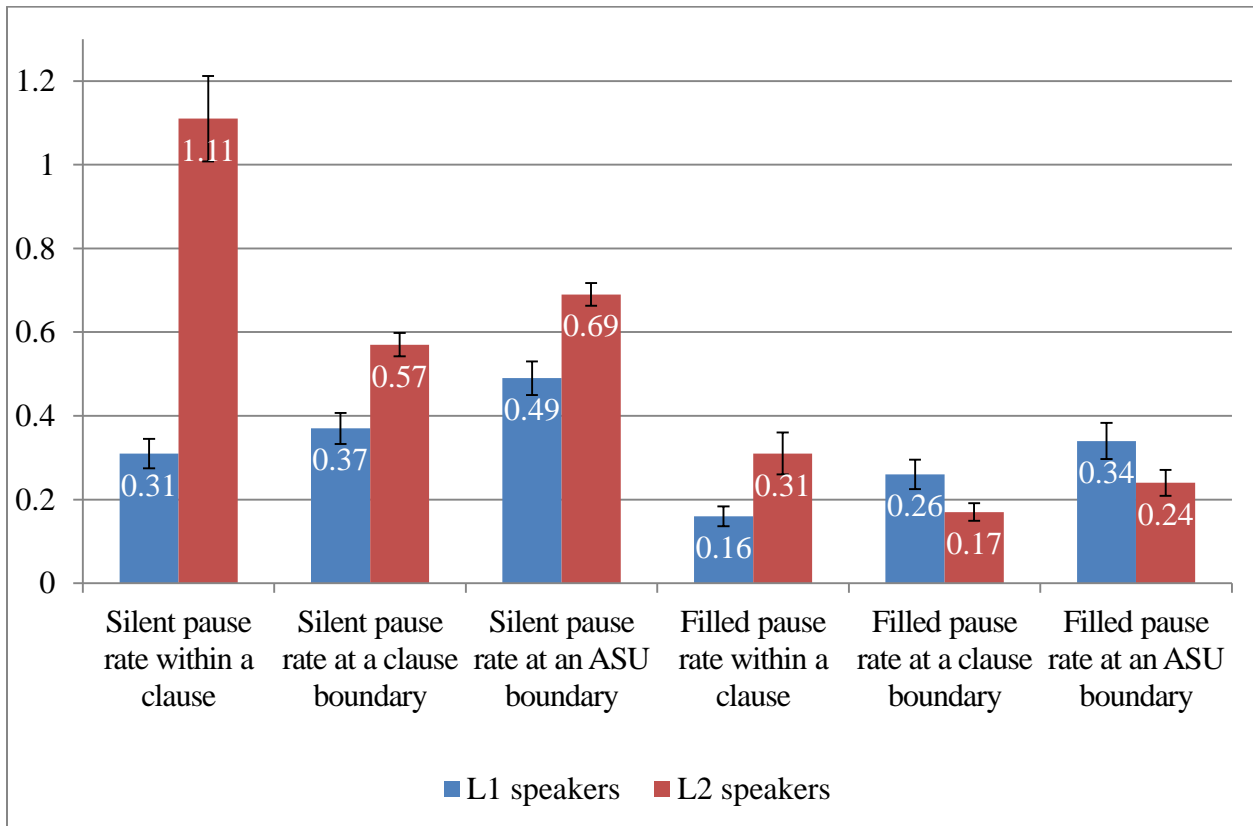
Figure 2: Distribution of speaking and pause time



In addition, as filled pauses were often preceded or followed by silent pauses, *Pause concurrence rate* was calculated (see Table 1) to find out how often they co-occurred in the speech of the two speaker groups. In L1 speech more than half of the pauses (53%) were adjacent to other pauses, whereas in L2 speech only 35% were.

Figure 3 illustrates the L1 and L2 speakers' *Pause rate within a clause, at a clause boundary, and at an ASU boundary* for silent and filled pauses.

Figure 3: Pause rates for silent and filled pauses in different locations (means and standard errors)



Pause rate in different locations demonstrates how often a speaker paused within each clause, at each clause boundary, and each ASU boundary (see Table 1) and 1 means the speaker paused in a given location every time. The L1 and L2 speakers demonstrated a different pattern in *Pause rate* regarding locations. The L1 speakers' *Pause rate* was the lowest within a clause, increased a bit at a clause boundary, and was the highest at an ASU boundary for both silent (0.31, 0.37, and 0.49) and filled pauses (0.16, 0.26, and 0.34). On the other hand, the L2 speakers'

Pause rate was the highest within a clause both for silent (1.11) and filled pauses (0.31). *Silent pause rate within a clause* especially yielded the most striking group difference, in which the rate for the L2 speakers to pause was almost four times higher than that for the L1 speakers (1.11 vs. 0.31). Moreover, the L2 speakers' *Silent pause rate within a clause* was over 1, suggesting that on average they paused more than once within each clause.

A one-way MANOVA showed that using Pillai's trace, the two groups were significantly different in overall *Pause rate*, $V = .60$, $F(6, 39) = 9.53$, $p < .001$. Univariate ANOVAs on the six dependent variables revealed significant group differences on *Pause rate* for silent pauses in all three locations (within a clause, $F(1, 44) = 28.37$, $p < .001$, $\eta_p^2 = .39$; at a clause boundary, $F(1, 44) = 16.14$, $p < .001$, $\eta_p^2 = .27$; at an ASU boundary, $F(1, 44) = 18.42$, $p < .001$, $\eta_p^2 = .30$) but not on *Pause rate* for filled pauses in any locations (within a clause, $F(1, 44) = 4.04$, $p = .05$, $\eta_p^2 = .08$; at a clause boundary, $F(1, 44) = 4.99$, $p = .03$, $\eta_p^2 = .10$; at an ASU boundary, $F(1, 44) = 4.20$, $p = .05$, $\eta_p^2 = .09$) at the .008 level after Bonferroni correction.

Table 3 shows the two speaker groups' mean length of silent and filled pauses in different locations.

Table 3: Mean length of silent and filled pause in different locations (ms)

	L1 Speakers ($N = 15$)		L2 Speakers ($N = 31$)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Mean length of silent pauses within a clause	621	179	811	233
at a clause boundary	722	198	1018	434
at an ASU boundary	734	218	1052	475
Mean length of filled pauses within a clause	498	119	488	192
at a clause boundary	503	126	588	247
at an ASU boundary	518	125	606	257

Note. Silent pause ≥ 250 ms.

The L1 and L2 speakers demonstrated a similar pattern in the mean length of pauses in terms of locations. For both speakers the length of pause was shortest within a clause and longest at an ASU boundary. A one-way MANOVA showed that using Pillai's trace, there was no significant difference between the L1 and L2 groups, $V = .22$, $F(6, 39) = 9.53$, $p = .13$.

Correlation Between the Temporal Variables and Speaking Scores

In order to investigate relationships between temporal variables and speaking scores, a Pearson correlation analysis was run with the L2 learners' data on speed, length of run, repairs, pause phenomena, and speaking scores (Table 4). In the analysis, the length of silent and filled pause in different locations is not included as there was no group difference in the measures according to the MANOVA test in the previous section.

First, to examine the relationships between the fluency measures, *Mean syllable duration* was positively correlated with *Mean syllables per run* ($r = .582^{**}$), *Mean length of filled pauses* ($r = .425^*$), and *Silent pause rate within a clause* ($r = .580^{**}$). *Mean syllables per run* was correlated with *Number of repetitions per minute* ($r = .461^*$) and with some silent pause measures. Considering that the calculation of *Mean syllables per run* involves number of silent pauses, it is not surprising to see the correlations between them; however, it is noteworthy that *Mean syllables per run* demonstrated the highest correlation with *Silent pause rate within a clause* ($r = -.855^{**}$) and a much lower correlation with *Silent pause rate at an ASU boundary* ($r = -.385^*$). *Number of corrections per minute* was not correlated with *Number of repetitions per minute* or the rest of the measures. On the other hand, *Number of repetitions per minute* was correlated with *Number of silent pauses per minute* ($r = .615^{**}$), *Mean length of filled pauses* ($r = -.412^*$), and *Silent pause rate within a clause* ($r = .581^{**}$), although they were not

mathematically related. *Number of filled pauses per minute* was negatively correlated with *Mean length of silent pauses* ($r = -.416^*$), which is consistent with the observation that filled pauses tend to interrupt the silence and break the silence into smaller pieces.

It is interesting to note that *Silent pause rate within a clause* had a moderately strong correlation with *Mean syllable duration* ($r = .580^{**}$) and *Number of repetitions* ($r = .581^{**}$) even though they were not mathematically related. It is also notable that *Pause rate at an ASU boundary* was very strongly correlated with that at a clause boundary ($r = .798^{**}$), whereas it was not correlated with that within a clause ($r = .134$), suggesting pauses within a clause and at an ASU boundary have a different pattern.

Finally, speaking scores had a moderately strong correlation with *Mean syllable duration* ($r = -.541^{**}$) and *Mean syllables per run* ($r = .549^{**}$). They were weakly correlated with *Number of repetitions per minute*, showing an approaching significance ($r = -.311$, $p = .089$) but not correlated with corrections ($r = -.048$, $p = .797$). Speaking scores were also not significantly correlated with *Number of silent pauses per minute* ($r = -.283$, $p = .123$) but weakly correlated with *Mean length of silent pauses*, showing an approaching significance ($r = -.341$, $p = .061$). They demonstrated a moderately strong correlation with *Silent pause rate within a clause* ($r = -.535^{**}$) and a weak correlation with *Silent pause rate at a clause boundary* ($r = -.358^*$), but no correlation with *Silent pause rate at an ASU boundary* ($r = -.017$). None of the measures on filled pause was correlated with speaking scores. It is noteworthy that the results of MANOVAs and ANOVAs in the previous section showed group differences in *Number of corrections per minute*, *Number of silent pauses per minute*, and *Silent pause rate at an ASU boundary*; however, these variables were not significantly correlated with speaking scores.

Table 4: Pearson correlations between utterance fluency measures and speaking scores

	MSD	MSR	Cor	Rep	SPmin	FPmin	LngSP	LngFP	SPRw	SPRc	SPRas	FPRw	FPRc	FPRas	Speak
MSD	1														-.541**
MSR	-.582**	1													.549**
Cor	.115	.043	1												-.048
Rep	.209	-.461**	.177	1											-.311
SPmin	.231	-.782**	.046	.615**	1										-.283
FPmin	.050	.212	.111	.099	-.023	1									.230
LngSP	.117	-.300	-.212	-.275	-.242	-.416*	1								-.341
LngFP	.425*	.168	.005	-.412*	-.345	.151	.009	1							.062
SPRw	.580**	-.855**	.227	.581**	.776**	-.159	.049	-.256	1						-.535**
SPRc	.276	-.618**	-.027	.152	.424*	-.255	.437*	.014	.433*	1					-.358*
SPRas	.072	-.385*	-.118	.023	.256	-.163	.328	.011	.134	.798**	1				-.017
FPRw	.187	.002	.088	.310	.130	.829**	-.287	-.015	.117	-.076	-.081	1			.050
FPRc	.281	-.228	-.129	-.020	.221	.503**	-.110	.237	.124	.092	.192	.196	1		.015
FPRas	.156	-.118	-.151	-.156	.117	.427*	-.096	.238	-.037	.071	.247	.052	.946**	1	.107

Note. * = $p < .05$; ** = $p < .01$. MSD = mean syllable duration; MSR = mean syllables per run; Cor = number of corrections per minute; Rep = number of repetitions per minute; SPmin = number of silent pauses per minute; FPmin = number of filled pauses per minute; LngSP = mean length of silent pauses; LngFP = mean length of filled pauses; SPRw = silent pause rate within a clause; SPRc = silent pause rate at a clause boundary; SPRas = silent pause rate at an ASU boundary; FPRw = filled pause rate within a clause; FPRc = filled pause rate at a clause boundary; FPRas = filled pause rate at an ASU boundary; Speak = speaking scores.

To summarize the results of utterance fluency, *Mean syllable duration*, *Mean syllables per run*, and *Silent pause rate within a clause* were most strongly associated with L2 oral fluency. The three measures not only distinguished between the L1 and L2 speakers with large effect sizes but also strongly correlated with L2 speaking scores. On the other hand, *Mean length of filled pauses* and *Mean length of pauses* in different locations (i.e., within a clause, at a clause boundary, and at an ASU boundary) did not distinguish between the L1 and L2 speakers.

Cognitive Fluency: Qualitative Study

Analysis

17 L2 speakers participated in the stimulated recall session right after the spontaneous speech task. During the session, they were asked to talk about what they were thinking while they were pausing or hesitating as far as they could remember. Stimulated recall was conducted in their L1, Korean, and their responses were translated into English by the Korean-English bilingual researcher. In order to investigate stimulated recall in terms of proficiency levels, the responses have been analyzed and reported by dividing the participants into two groups based on their speaking scores; 9 participants in the lower ($M = 40$, $SD = 5$ on a scale of 20 – 60) and the other 8 in the higher proficiency group ($M = 54$, $SD = 5.4$ on a scale of 20 – 60). All 8 participants in the higher proficiency group were undergraduate or graduate students at a university in the United States and had a minimum score of 90 on the internet-based TOEFL. On the other hand, the 9 participants in the lower proficiency group were either students in English as a second language (ESL) courses or university students who were still taking ESL courses. The mean length of residence in English speaking countries for the lower proficiency group was 0.7 years and for the higher proficiency group was 2.2 years.

The responses were categorized based on their common themes. Five main categories emerged: content of the message, vocabulary, grammar, phonology, and pragmatics. Grammar included responses about the issues on morphological and syntactic aspects, and the use of function words such as articles and prepositions. The examples of each category are presented and discussed in the results section. The data should be interpreted with caution in that stimulated recall has a limitation in capturing cognitive processes accurately, and it is not always straightforward to match a response with a specific pause or aspect of speech.

Results

Overall, the lower proficiency (LP) group reported 122 issues and the higher proficiency (HP) group reported only 75 issues. On average, the LP group also responded longer ($M = 13.2$, $SD = 3.9$ minutes) than the HP group did ($M = 9.2$, $SD = 1.8$ minutes). Furthermore, only 20% of the responses were marked with overt repairs in their speech samples, suggesting that 80% of the issues would have been missed through speech analysis without the stimulated recall data.

Table 5 shows that the comments on the content of the message comprised the largest proportion of comments by both groups (LP learners: 33.6%, HP learners: 50.7%). The LP learners mentioned grammar and vocabulary much more often than the HP learners, and the comments on grammar comprised one third of their responses (31.1%). In the following section, stimulated recall responses are analyzed qualitatively with examples for each category. The examples are to demonstrate some of the most common types of responses and to compare responses from the LP with the HP learners. For each example, information is provided about the quartile in which the L2 participant's speaking score falls.

Table 5: Distribution of stimulated recall responses

	Lower Proficiency Learners (N = 9)		Higher Proficiency Learners (N = 8)	
	Number of issues	Percentage of issues	Number of issues	Percentage of issues
Content of the message (L2 related issues)	41 (15)	33.6	38 (2)	50.7
Vocabulary	23	18.9	9	12.0
Grammar	38	31.1	14	18.7
Phonology	3	2.5	2	2.7
Pragmatics	2	1.6	5	6.7
Others (Individual differences, L1 use, etc.)	15	12.3	7	9.3
Total	122	100	75	100

Content of the Message

Comments on content were mainly about what to say next; however, interestingly, 37% of LP learners' comments on content were related to how their L2 proficiency affected planning their message, whereas only 5% of HP learners' comments were. Example 1 illustrates an LP learner (second quartile) who selected what to say considering her L2 competence.

1. among the interests (453) my interest is juvenile delinquency // (1198) um512 (812) I like it // because (1075) I think :: in juvenile delinquency is (378) very important //

Retrospection: *“My major is important” is not the reason that I like my major. But I just said that because I cannot explain other reasons in English.*

LP learners also mentioned that they often dropped or modified their original message because of their limited L2 competence as shown below (first quartile).

2. when I was (320) athlete :: {I was} um300 I (366) liked :: (260) for watching TV or movie //

Retrospection: *My original message was “I liked TV but I was afraid of becoming addicted to it. So I tried not to watch TV a lot.” But this was too difficult for me. So I just used an easy sentence and said “I liked watching TV.”*

In Example 2, the speaker abandoned his original message and said something opposite to his original message because the original message was too complex for him to convey in English. Moreover, according to the LP learners, abandoning the original message created an additional challenge.

3. uh498 (1243) when I have a free time :: I usually (817) do make something (462) like uh155 food // (887) and (883) {I} (498) I know :: I am very (250) awkward (643) cook //

Retrospection: *After I decided to drop a large part of my original message such as painting my apartment and furniture, and to talk about cooking only, I realized that I had to think of what to say all over again.*

The LP learner in Example 3 (second quartile) dropped part of his original message due to his limited L2 competence; however, this resulted in another challenge requiring him to plan a new message. It shows how an issue at one stage can interact with processing at another stage.

Vocabulary

LP learners commented on looking for words or deciding on expressions more often than the HP learners. Examples 4 and 5 are comments from an LP (first quartile) and an HP learner (fourth quartile), respectively.

4. I um594 (529) accept my mom {advise} advice //

Retrospection: *I was asking myself “what is ‘padadeulida’ (‘accept’ in Korean) in English?”*

5. if we (809) ah174 create very valid (538) uh421 (500) assessment tools :: then I think :: the more students can benefit (658) ah272 from that (737) ah414 assessment system //

Retrospection: *I was thinking whether I would say “valid test” or “valid assessment tool.”*

As exemplified above, although both comments concerned vocabulary, there seem to be qualitative differences between them. In 4, the LP learner was trying to look for an L2 word by translating it from her L1, whereas the HP learner in 5 already had two multi-word expressions retrieved and was trying to decide which one to use to better fit the context. Furthermore, LP learners sometimes tried to decide which vocabulary to use by accessing L2 declarative knowledge regarding word choice as below (first quartile).

6. I enjoy :: (1504) um1090 {(in a whisper) watching seeing looking (288) reading} (333) reading books //

Retrospection: *I got confused among watching, seeing, looking, and reading. I learned that with “books”, I should say “reading.”*

Grammar

Comments on grammar by the two groups also had not only quantitative differences but also qualitative differences. The LP learners remembered their problems regarding sentence construction and choice of tense-aspect, articles, and prepositions. They reported they were often thinking about specific rules, L2 declarative knowledge they had learned, for instance, how to make a sentence using comparatives, whether to use a gerund or an infinitive. On the other hand, the majority of HP learners' comments on grammar involved monitoring their sentence structure, especially with complex clauses. Examples 7 and 8 illustrate LP learners (first quartile) having

difficulty deciding on tense-aspect and making a comparative sentence using their L2 declarative knowledge.

7. uh468 I learned (450) uh699 textile (1312) // and (490) I don't (267) like my major (283) // because it's not match for (357) me //

Retrospection: *I was wondering whether I should use the present tense or present perfect because I didn't like it in the past and I still don't like it.*

8. In Korea th (1853) {golf} (1335) golf cost is very expensive // (1814) but (463) I surprised :: that (856) uh268 in USA (440) very cheap than {in my} in my country //

Retrospection: *Instead of just saying that it is expensive to play golf in Korea, I wanted to make a sentence which compares the golf cost between Korea and the US and say it is more expensive to play golf in Korea than in the US but it wasn't easy to put the sentence together.*

Phonology and Pragmatics

Only a few comments were made on phonology and pragmatics by both groups, although the HP group commented on pragmatics more frequently than the LP group. Most comments on pragmatics concerned repetition of words or expressions. Example 9 illustrates an LP learner (second quartile) monitoring her own pronunciation.

9. ts (476) {is criminal justice} in criminal justice there are (329) lots of academic interests // (1176) for example (784) murder rapes (862) and thief (521) perjury // (1058) uh711 (358) any kinds of crime // (873) Ts (526) and {in addish} in addition (358){ju} (358) juvenile delinquency (633)

Retrospection: *When I said “in addition,” I thought my intonation was wrong, so I tried to repair stress and intonation. As “juvenile” is such a difficult word for me to pronounce, I always monitor whether I pronounce it correctly.*

Other Issues

Responses also showed that individual differences might play a role in deciding whether to repair one’s errors and influence oral fluency. Following are the comments from two LP learners (second quartile) on their decision regarding self-repair.

10. *...some people just disregard their mistakes even after they notice them, but I don’t like to ignore them. So when I make mistakes, I try to repair them, even if this often gets my sentences all mixed up...*

11. *...before I came to the US, I thought about how to express things in English carefully before speaking but these days I just try to say things without thinking too much. Because after I came here I realized that overthinking didn’t seem to really improve my utterance and when I make errors in English, some English speakers like my teacher also help me to repair them.*

In 10 the learner seems to attribute her tendency to repair her mistakes to her personality and the learner in 11 explains how his L2 learning experience changed his self-repair behavior.

Other issues that the L2 learners mentioned included difficulty connecting speech messages due to constant monitoring and access to L2 declarative knowledge, use of L1 in conceptual preparation, and translating messages from L1 to L2. A few higher proficiency learners reported difficulty remembering what they were thinking while pausing, saying “I don’t remember.”

Discussion

The current study aimed to demonstrate in what respects L1 and L2 speakers' fluency are different. Within Segalowitz's (2010) fluency framework, the study investigated (1) utterance fluency by comparing speech samples from L1 and L2 speakers, and (2) cognitive fluency by examining the stimulated recall responses from L2 speakers.

Utterance Fluency

Utterance fluency was investigated in terms of speed, length of run, repairs and pause phenomena. Table 6 summarizes the results of the MANOVAs and correlation analysis. Based on the results, *Mean syllable duration*, *Mean syllables per run*, and *Silent pause rate within a clause* were most strongly associated with L2 oral fluency. The three measures not only distinguished between the L1 and L2 speakers with large effect sizes but also significantly correlated with L2 speaking scores. On the other hand, *Mean length of filled pauses* and *Mean length of pauses* in different locations (i.e., within a clause, at a clause boundary, and at an ASU boundary) did not distinguish between the L1 and L2 speakers. In the following, the findings on speed, length of run, repairs, and pause phenomena are discussed in turn in relation to previous studies.

Speed

In the current study, *Mean syllable duration*, which excludes pause time and is thus a pure measure of speed (De Jong, Steinel, et al., 2013), distinguished the L1 and the L2 speakers and exhibited a significant negative correlation with L2 speaking scores ($r = -.541$).

Table 6: Summary of statistical analyses

Measures	Submeasures	L1-L2 difference	Correlation with speaking scores
Speed	Mean syllable duration (ms)	✓	-.541**
Length of run	Mean syllables per run	✓	.549**
Repair	Number of corrections per minute	✓	-.048
	repetitions per minute	approaching	-.311
Pause	Number of silent pauses per minute	✓	-.283
	filled pauses per minute	approaching	.230
	Mean length of silent pauses (ms)	approaching	-.341
	filled pauses (ms)		.062
	Silent pause rate within a clause	✓	-.535**
	at a clause boundary	✓	-.358*
	at an ASU boundary	✓	-.017
	Filled pause rate within a clause		.050
	at a clause boundary		.015
	at an ASU boundary		.107
	Mean length of silent pauses within a clause		
	at a clause boundary		
	at an ASU boundary		
	Mean length of filled pauses within a clause		
	at a clause boundary		
	at an ASU boundary		

Note. ✓ = statistical difference after Bonferroni corrections (for details, see the analysis section). * $p < .05$; ** $p < .01$. As results of MANOVA showed that there was no group difference in mean length of pauses in different locations (the bottom 6 rows), the variables were not included in the correlation analysis.

Previous studies on perceived fluency often did not find a correlation between *Articulation rate* (comparable to *Mean syllable duration*) and fluency ratings. In Kormos and Dénes (2004) and Cucchiarini et al. (2002) *Articulation rate* was not correlated with fluency ratings. On the other hand, results of the present study are consistent with recent studies which investigated relationships between acoustic fluency measures and oral proficiency or cognitive fluency. In Ginther et al. (2010), *Articulation rate* was significantly correlated with oral proficiency scores ($r = .61$). In De Jong, Steinel, et al. (2013), *Mean syllable duration* explained 50% of the variance of linguistic knowledge and skills (cognitive fluency). De Jong, Groenhout,

et al. (2013) also found that L2 *Mean syllable duration* explained 30% and the corrected measure for L1 behavior explained 41% of variance of L2 proficiency. Therefore, as suggested by De Jong, Steinel, et al. (2013), pure measures of speed such as *Mean syllable duration* and *Articulation rate* can be claimed to reflect L2 cognitive fluency and L2 proficiency.

Length of Run

As in many previous studies (e.g., Kormos & Dénes, 2004; Lennon, 1990; O'Brien et al., 2007; Segalowitz & Freed, 2004; Towell et al., 1996), *Mean syllables per run* was strongly associated with L2 fluency by demonstrating a large difference between the L1 and L2 speakers and a significant correlation with speaking scores. Length of run has a conceptual connection with automatic speech production processing. Towell et al. (1996) claimed that increase in length of run reflects proceduralization of declarative knowledge based on Anderson's (1983) ACT* model of skill acquisition. Furthermore, length of run also seems to be closely related to the use of prefabricated language units and formulaic language, which have been claimed to facilitate L2 oral fluency (Boers et al., 2006; Bybee, 2002; Kuiper, 1996; Skehan, 1998). However, the direct connection between automaticity and length of run is still not clear. For instance, it is unclear whether long fluent run reflects automaticity in lemma retrieval, grammatical encoding, morpho-phonological encoding, phonetic encoding, articulation, or combination of some of these stages. Segalowitz and Freed (2004) found correlations between mean length of run without fillers and lexical access speed and efficiency, suggesting its potential connection with lemma retrieval; however, whether length of run is also related to other processing stages such as grammatical encoding or morpho-phonological encoding requires further research.

Repairs

Repair measures have often not been indicative of perceived fluency in previous studies (e.g., Cucchiarini et al., 2002; Kormos & Dénes, 2004). However, Bosker et al. (2013) found that repairs did contribute a small but significant amount to perceived fluency. In the present study, the L2 speakers used more corrections and repetitions than the L1 speakers, and there was a weak negative correlation between repetitions and speaking scores. The data also suggest they are affected by individual differences, as demonstrated by the large standard deviations of the frequency of repairs in L1 as well as L2 (Table 2), and also by the stimulated recall data (Examples 10 and 11), in which personality and L2 learning experience played a role in deciding whether or not to repair one's errors.

Another issue regarding repairs is that self-corrections and repetitions exhibited different relations with L2 fluency. In the correlation analysis, number of corrections and repetitions did not correlate with each other. Moreover, only repetitions were correlated with speaking scores (showing an approaching significance) and corrections were not. Number of repetitions was also correlated with number of silent pauses, length of filled pauses, and silent pause rate within a clause but number of corrections was not. The differential findings between the number of corrections and repetitions are in line with De Jong, Steinel, et al. (2013), in which number of corrections explained 25% of the variance but number of repetitions explained 12% of the variance of linguistic knowledge and skills (cognitive fluency). Therefore, it would be premature to claim that repairs do not reflect L2 fluency and future research on repairs will need to address the issues regarding individual differences and potentially differential roles of self-corrections and repetitions in L2 fluency.

Pause Phenomena

As mixed results had been found in the literature on breakdown fluency, the present study especially analyzed pause phenomena rigorously by investigating frequency, duration, and distribution of both silent and filled pauses. In this section, results on pause phenomena are discussed in depth in relation to previous studies.

Results on pause phenomena showed that the L1 and L2 groups differed more in the frequency than length of pauses and they were also more different in the use of silent pauses than filled pauses. The results are consistent with the findings in the literature that pause duration was often not strongly associated with L2 fluency. It was not correlated with fluency ratings (e.g., Cucchiarini et al., 2002) and only very weakly related to cognitive fluency (De Jong, Steinel, et al., 2013). De Jong, Groenhout, et al. (2013) further showed that length of silent pauses barely explained the variance of L2 proficiency (0 – 3%). In her comparative research on L1 and L2 pausing, Riazantseva (2001) concluded that pause duration is a language-specific feature.

In terms of filled pauses, it is interesting to note that the L2 speakers used filled pauses less often than the L1 speakers, showing an approaching significance, and L2 speakers' filled-to-silent-pauses ratio was more than two times lower than L1 speakers'. Filled pauses can be viewed either as a non-linguistic symptom for trouble in the speech production process (Goffman, 1981; Levelt, 1989), or as “fillers” with linguistic elements such as discourse markers “well” and “you know.” Clark and Fox Tree (2002) proposed “uh” and “um” are words with functions. From the perspective of their “filler-as-word hypothesis,” appropriate use of filled pauses can be considered as something that L2 learners should acquire and have a relationship with L2 proficiency. In fact, the correlation analysis showed that speaking scores had a weak but *positive* correlation with *Number of filled pauses per minute* ($r = .230$), while they had a weak *negative*

correlation with *Number of silent pauses per minute* ($r = -.283$). Given that the L1 speakers also had a higher concurrence rate of silent and filled pauses compared to the L2 speakers, L1 speakers may use filled pauses strategically to signal a delay in speaking and splice a long silent pause into smaller pauses, making their speech sound less disfluent. In the current study, *Number of filled pauses per minute* did negatively correlate with *Mean length of silent pauses* ($r = -.416$), which is consistent with the observation that filled pauses tend to interrupt the silence and break the silence into smaller pieces. The relationship between filled pauses and L2 proficiency also needs further investigation.

Although the MANOVA results showed some group differences in the frequency and length of pauses, it is noteworthy that none of the overall pause measures was significantly correlated with speaking scores until pause location was taken into account. *Silent pause rate within a clause* had the most striking group difference (Figure 3) and a strong negative correlation with speaking scores⁵. The findings are corroborated by previous studies (e.g., Deschamps, 1980; Tavakoli, 2011), in which L2 speech often had pauses in the middle of clauses, whereas L1 speech had pauses at syntactic boundaries. The significant correlation between the speaking scores and *Silent pause rate within a clause* is compatible with the findings on L2 fluency development (Lennon, 1990; Towell et al., 1996) and the argument that pauses within clauses reflect processing difficulties in speech production (e.g., Lennon, 1984; Pawley & Syder, 2000; Wood, 2010).

Research on L1 speech production has shown that pauses within clauses typically occur before unpredictable and infrequent words and are related to lexical retrieval (Levelt, 1983;

⁵ These findings can be interpreted with even more emphasis because when speakers make longer clauses (e.g., L1 speakers, higher proficiency learners), they have a higher chance of pausing within a clause by definition.

Maclay & Osgood, 1959), whereas pauses at clause boundaries are associated with a more general “long-term” planning of the following clause such as word ordering and syntactic encoding (Kircher, Brammer, Levelt, Bartel, & McGuire, 2004). Kircher et al. (2004) used functional Magnetic Resonance Imaging (fMRI) to examine neural correlates of pauses within clauses and found that pauses within clauses are associated with left temporal activation. They claimed that their findings suggest that pauses within clauses are a correlate of speech planning and in particular lexical retrieval. Based on their claim, L2 speakers’ high *Silent pause rate within a clause* in the current data can be interpreted to reflect difficulty in speech planning and lexical retrieval. The interpretation is also consistent with the observation that lexical retrieval is one of the most salient problems with L2 speakers (De Bot, 1992). Furthermore, as shown in the stimulated recall data and their speech in the previous section, unlike L1 speakers, L2 speakers may pause in the middle of a clause even for more general planning such as deciding content of message (Examples 1 and 2) and syntactic encoding (Example 8) because their speech planning processes are not automatized and often go through a process of trial and error during speech production. These speech planning processes are likely to result in high pause rate within clauses by the L2 speakers.

Moreover, the measure for the distribution of pauses, *Pause rate* is worth mentioning. *Pause rate* was devised in the current study to measure how likely a speaker pauses within each clause, at each clause boundary, and at each ASU boundary by dividing number of pauses by corresponding unit (i.e., number of clauses, clause boundaries, ASU boundaries). *Pause rate* is able to capture the frequency of pauses in different locations more accurately than measures used previously such as *Number of pauses per minute* (e.g., Tavakoli, 2011). In Table 7, *Number of pauses per minute* is calculated. It shows that the L1 speakers had more pauses at clause

boundaries than at ASU boundaries; however, the results are not very informative and make it hard to compare pause frequency across the locations because there were more clauses ($M = 31$) than ASUs ($M = 18$). In fact, when number of clauses and ASUs were controlled, results of *Silent pause rate* (Figure 3) showed that the L1 speakers paused at an ASU boundary (0.49) more often than within a clause (0.31) or at a clause boundary (0.37).

Table 7: Number of silent and filled pauses per minute in different locations

	L1 Speakers ($N = 15$)		L2 Speakers ($N = 31$)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Number of silent pauses per minute				
Within clauses	6.6	2.2	13.8	4.5
At clause boundaries	8.5	2.1	8.0	2.3
At ASU boundaries	6.8	1.5	7.0	2.1
Number of filled pauses per minute				
Within clauses	3.9	2.0	4.2	3.7
At clause boundaries	5.5	2.5	2.3	1.8
At ASU boundaries	4.5	2.0	2.2	1.8

Pause rate also enables accurate comparison of pause distribution across speakers.

According to Table 7, the L1 and L2 speakers were similar in *Number of silent pauses* at clause boundaries or at ASU boundaries; however, again L1 speech had more clauses and ASUs (31 clauses; 18 ASUs) than L2 speech (24 clauses; 16 ASUs). After controlling for the number of clauses and ASUs across the speakers, results of *Pause rate* show that the L2 speakers also had a higher pause rate than the L1 speakers at a clause boundary and at an ASU boundary (Figure 3). Based on L2 speakers' high *Silent pause rate within a clause*, one might have predicted that, unlike within a clause, at clause or ASU boundaries L2 speakers would pause less often than L1 speakers. *Pause rate* was also able to precisely address this question and shows that, contrary to

the prediction, the L2 speakers, in fact, paused more often than the L1 speakers not only within a clause but also at both boundaries even if the group difference was much more striking within a clause. Given that L1 speakers' pauses between clauses are associated with general speech planning such as word ordering and syntactic encoding (Kircher et al., 2004), it seems logical for L2 speakers to pause between clauses as well as within clauses more often than L1 speakers.

Cognitive Fluency

In investigating L2 cognitive processes and issues regarding disfluencies, the current study used stimulated recall which can reflect cognitive events and reveal the information attended to during task performance (Gass & Mackey, 2000) and this technique has rarely been used in studies of L2 fluency. The data showed that only 20% of the responses were marked with overt repairs in speech samples, suggesting 80% of the issues would have been missed through speech analysis without stimulated recall.

According to the declarative/procedural model (Ullman, 2001, 2004, 2005) and the studies on automaticity (DeKeyser, 2001, 2007; Segalowitz, 2000, 2003), practice and more exposure to L2 lead to dependence on procedural memory/knowledge. Based on their claims, lower proficiency learners were predicted to rely on L2 declarative knowledge more than higher proficiency learners. Considering that only declarative knowledge can be explicitly recollected (Ullman, 2001, 2004, 2013), lower proficiency learners were expected to remember more about their thoughts at the time of speaking than higher proficiency learners. Results followed the predictions and on average the lower proficiency learners reported over 1.5 times more issues than the higher proficiency learners did and also responded 44% longer than the higher proficiency learners. It may not be very surprising that the lower proficiency learners commented

on more cases than the higher proficiency learners as they reported what they were thinking while pausing or hesitating at the time of speaking. As the lower proficiency learners paused or hesitated more often than the higher proficiency learners, they had more pauses to talk about. However, it is still interesting to see that whereas the lower proficiency learners remembered and reported a number of their thoughts quite easily, the higher proficiency learners often mentioned having trouble remembering them, saying that “I don’t remember.” The findings are also in line with the argument that lower proficiency learners consciously think about more issues during the speech production process because the process has not yet been automatized and requires a lot of attentional effort (Kormos, 2006).

Furthermore, as higher proficiency learners’ production processes are assumed to be more automatized than lower proficiency learners’, the higher proficiency learners were predicted to report mainly on macroplanning and monitoring that required their attention as in L1 speech (Kormos, 2006), whereas the lower proficiency learners were expected to report on more varied issues including syntactic and morpho-phonological encoding that are not fully automatized, and controlled in the declarative memory system. The predictions were met and the content of the message comprised the larger proportion of the comments by the higher proficiency learners (51%) than the lower proficiency learners (34%), and the lower proficiency learners commented on L2 declarative knowledge concerning grammar and vocabulary much more often than the higher proficiency learners.

Some part of the stimulated recall responses can be discussed in association with the L2 speech production model and the fluency vulnerability points in Figure 1. In both L2 groups’ responses, the largest proportion consisted of the content of the message. Considering that even L1 speakers’ fluency declines during macroplanning (Roberts & Kirsner, 2000), it seems

reasonable that L2 speakers pause for a considerable amount of time to think about what to say. However, more importantly, 37% of lower proficiency learners' comments on content were related to their limited L2 competence, whereas only 5% of higher proficiency learners' were. The lower proficiency learners often dropped or modified their original message due to L2 difficulties (Example 2), as reported in the studies on monitoring and problem-solving mechanisms (Dörnyei & Kormos, 1998; Kormos, 2000a, 2006). It is also interesting to note that the lower proficiency learners often abandoned their original message to avoid problematic situations; this eventually imposed another cognitive burden that they had to plan a new message (Example 3).

Furthermore, the responses showed that the L2 speakers were affected by their L2 proficiency not only when monitoring their message after articulation, but also at an early stage of conceptual preparation as in Example 1, in which an L2 learner chose what to say considering her L2 competence. Following is a comment from a higher proficiency learner, which precisely addresses the issue focusing on the proficiency level.

12. *...when I choose what to say, I think my English proficiency influences the decision. I think that "I can say this much, so I will say this." When I was a beginner in English, there were so many cases in which I failed to convey what I originally planned to say. So now I decide what to say based on how much I can express in English...*

The comment suggests that as proficiency levels increase, L2 learners not only develop L2 competence and skills, but may also become more strategic about deciding what to say possibly based on their understanding of their own oral proficiency through L2 experience. It is possible that lower proficiency learners do not have a good understanding of their own oral proficiency, choose a relatively difficult or complex topic for them to talk about, face difficulties

in conveying the original message, decide to drop the message, and face another challenge to pick a better topic for them to talk about. The trial and errors are likely to be associated with more pauses and repairs, interrupting speech fluency. This is compatible with Segalowitz's (2010) prediction that "the more macroplanning a communicative situation requires, the more vulnerable the L2 speech will be to disfluencies because of the diversion of processing resources" (p.11). De Bot (1992) also suggested that non-balanced bilinguals sometimes do not have the lexical items required to express a concept and they often seem to be aware of this problem in advance and take it into account in conceptual preparation. In their studies on problem-solving mechanisms and self-monitoring in L2 speech, Dörnyei and Kormos (1998) and Kormos (2000a) also exemplify a number of cases when L2 speakers abandon or change their original intended message (macro-plan) due to L2 difficulties. As the L2 proficiency level and learners' level of understanding about their own oral proficiency can play a role in deciding what to say, the current study proposes that macroplanning be another candidate for a fluency vulnerability point in the L2 speech production model (Figure 1; Segalowitz, 2010).

Another point to discuss in relation to the model of L2 speakers concerns the interpretation of the finding that the lower proficiency learners remembered attending to L2 declarative knowledge regarding specific grammatical features and vocabulary much more often than the higher proficiency learners. This might be considered to reflect processing difficulties in grammatical encoding and lemma retrieval in the model. However, there are a number of issues to be resolved to pinpoint at which speech production stage these difficulties occurred. For instance, a number of responses on grammar by the lower proficiency learners concerned the choice of tense-aspect. They often mentioned that they tried to apply what they had learned in class. However, the information about tense-aspect and mood is claimed to be added during

microplanning in the L1 speaking model through an automatic process without requiring attentional effort (Levelt, 1989). Here we face a number of L2 specific issues which the L1 speech production model does not seem to address. For instance, in L2 it is not clear whether the information about tense-aspect is still added to a preverbal message during microplanning. If the microplanning stage itself is highly automatic in nature, it is not easy to explain why so many L2 learners were able to remember thinking about tense-aspect.

In fact, all the boxes in the model (Figure 1) represent processing components with procedural knowledge, thus being largely automatic and subconscious in nature. By contrast, based on the responses, the lower proficiency learners almost always seemed to think about L2 declarative knowledge or rules while speaking in L2 (e.g., deciding on the words, tense-aspect, function words such as articles, prepositions, and putting words together to construct a sentence). This seems to support Kormos' (2006) proposal to add an L2 declarative knowledge storage in the L2 speech production model⁶. However, it is still not clear which processing stages of L2 speech production have access to L2 declarative knowledge or how the L2 declarative storage and other speech production stages interact. There are a number of issues to be investigated in order to understand the process of L2 speech production.

General Discussion: Study 1

⁶ It is true that stimulated recall data had a number of comments on L2 declarative knowledge because stimulated recall taps the conscious, thus mainly declarative processes. However, what this study tried to show is not that subconscious processes are involved less, or are less important in L2 speech production than in L1 speech production, but rather to show that conscious processes seem to be frequently involved in L2 production, which is not easy to explain based on the L1 speech production model.

To demonstrate in what respects L1 and L2 speakers' fluency are different, Study 1 investigated utterance fluency and cognitive fluency of the L1 English speakers and the L1 Korean L2 English speakers by analyzing speed, length of run, repairs, and pause phenomena of their speech samples, and stimulated recall responses.

The results of the MANOVAs and correlation analysis showed that *Mean syllable duration*, *Mean syllables per run*, and *Silent pause rate within a clause* were most strongly associated with L2 oral fluency. The three measures not only distinguished between the L1 and L2 speakers with large effect sizes but also strongly correlated with L2 speaking scores. Although pure speed measures (i.e., *Articulation rate*, *Mean syllable duration*) have not always been found to be a strong associate of perceived fluency in the literature, the current finding is consistent with recent studies which showed a strong correlation between speed fluency with oral proficiency and cognitive fluency (Ginther et al., 2010; De Jong, Groenhout, et al., 2013; De Jong, Steinel, et al., 2013). Study 1 especially conducted an in-depth analysis on pause phenomena by examining frequency, duration, and distribution of both silent and filled pauses. The results showed that *Silent pause rate within a clause* clearly distinguished the two groups and had a strong correlation with speaking scores. The findings are consistent with the claims in previous studies that pauses within clauses reflect processing difficulties in speech production (e.g., Kircher et al., 2004; Pawley & Syder, 2000).

Stimulated recall responses showed that the lower proficiency learners remembered more issues regarding L2 declarative knowledge on grammar and vocabulary than the higher proficiency learners, which was compatible with the declarative/procedural model and studies on automaticity. In addition, stimulated recall responses suggested a possibility that macroplanning be another candidate for a fluency vulnerability point, considering that L2 proficiency seems to

affect L2 speakers' initial decision on the message. Lower proficiency learners' frequent comments on specific grammatical rules and vocabulary seem to lend support for including an L2 declarative knowledge store in the L2 speech production model.

The present study tried to fill gaps and extend the body of research on L2 fluency by investigating utterance and cognitive fluency within Segalowitz's (2010) framework. The study provided empirical evidence to demonstrate in what respects L1 and L2 speakers' fluency are different by examining temporal measures of speed, repairs, and pause phenomena. It analyzed utterance fluency in a comprehensive and rigorous way to address less studied aspects of fluency (e.g., distribution of pauses, frequency and duration of filled pauses, relationship between silent and filled pauses) and inconclusive findings in previous studies (e.g., pure measure of speed fluency, repairs).

Study 1 also has implications for research methodology. The study discussed strengths and weaknesses of different measures used in previous studies and further proposed a measure (i.e., *Pause rate*) that can depict pause distribution more accurately. It also utilized qualitative analysis in exploring cognitive fluency to provide additional insight to the field of L2 fluency research where quantitative analysis is dominant.

Furthermore, results of the study have potential implications for L2 education and assessment. One of the most novel and important findings of the current study is the close relationship between L2 utterance fluency and pauses within clauses. L1 and L2 speech exhibited a striking difference in the frequency of pauses within clauses, which is considered to reflect difficulties in speech production processing such as lexical retrieval. Based on the findings, in classroom one way teachers can help L2 learners to enhance L2 fluency is to provide ample opportunities to practice collocations and formulaic language, which will enable learners to

produce longer fluent runs and will decrease pauses within clauses in their speech. In terms of L2 assessment, including the measure which addresses the frequency of silent pauses within clauses in automatic fluency assessment can evaluate L2 learners' oral fluency more accurately.

CHAPTER 3: STUDY 2

Introduction

In order to identify speech features which affect the perception of fluency, a number of previous studies investigated the relationship between utterance fluency and perceived fluency by relating fluency ratings to acoustic characteristics of L2 speech (e.g., Bosker et al., 2013; Cucchiarini et al., 2000, 2002; Derwing et al., 2004; Freed, 2000; Ginther et al., 2010; Kormos & Dénes, 2004; Rossiter, 2009). However, results are still mixed regarding pause phenomena. For instance, pause frequency was correlated with fluency ratings in Rossiter (2009) but not in Kormos & Dénes (2004); whereas pause duration was correlated with fluency ratings in Kormos & Dénes (2004) but not in Cucchiarini et al. (2002). Although pause phenomena seem to play an important role in the perception of fluency, the relative contributions of frequency, duration, and distribution of pauses have rarely been investigated. In particular, it has been argued that pauses within constituents, which recent studies have identified as major characteristics of non-fluent L2 speech (Kahng, 2012; Tavakoli, 2011), reflect difficulties in speech processing and planning. However, the effects of pause location on perceived fluency of L2 speech have not yet been examined.

As discussed in the literature review section, L1 research on pause phenomena suggests that pause location affects speech perception. Silent pauses are one of the acoustic cues to clausal units along with pitch and vowel duration (Seidl & Cristià, 2008). Therefore, silent pauses at grammatical boundaries have been claimed to help listener comprehension by indicating the boundaries of speech to be analyzed, and by providing cognitive processing time (e.g., Arons, 1993; Griffiths, 1991; Reich, 1980, Sugito, 1990), whereas pauses within clauses can be

disrupting. It has also been reported that beneficial effects of silent pauses between clauses on listeners were apparent only under conditions of cognitive complexity in auditory speech processing and they did not demonstrate beneficial effects when the speech or tasks were easy enough (Aaronson, 1968; Reich, 1980).

Study 2 aims to address the gaps in the literature on L2 perceived fluency and examined 1) the relative contributions of frequency, duration, and distribution of pauses to the perception of L2 fluency (Experiment 1), and 2) the effects of pause locations on the perception of L1 and L2 fluency (Experiment 2).

Experiment 1

Experiment 1 investigated the relative contributions of frequency, duration, and distribution of silent pauses to fluency ratings. The research questions for Experiment 1 are:

1. Which acoustic measures of pause phenomena (frequency, duration and/or distribution of silent pauses) are significantly related to fluency ratings?
2. Does the distribution of pauses explain significantly additional variance of fluency ratings which is not explained by frequency and duration of silent pauses?

Based on previous studies on perceived fluency, frequency and duration of silent pauses are predicted to be correlated with fluency ratings. On the other hand, the relationship between the distribution of silent pauses and L2 perceived fluency has not been investigated; therefore, it is the main focus of Experiment 1. Research on L1 pause phenomena has shown that silent pauses are one of the cues to clausal units, and pauses between clauses can be useful, whereas

pauses within clauses can interfere with speech perception processing (Bower & Springston, 1970; Griffiths, 1991; Reich, 1980; Sugito, 1990). On the basis of the L1 research findings, fluency ratings are predicted to correlate with not only frequency and duration of silent pauses but also distribution of silent pauses. In addition, if the regression model with the variable of pause distribution explains significantly larger variance of fluency ratings than the model without the variable of pause distribution, the result can be interpreted to reflect its critical role of pause distribution in perceived fluency.

In Experiment 1, English native listeners rated L2 speech samples on fluency level. The speech samples were also acoustically analyzed in terms of frequency, duration, and distribution of silent pauses. The relative contributions of the three aspects of pause phenomena to fluency ratings were examined through multiple regression analysis.

Method

Raters

Forty-six native English speakers (16 male; 30 female) participated in the experiment as raters. They were undergraduate students at a large university in the United States and the mean age was 21 ($SD = 2.3$) and reported to have normal hearing. Their mean familiarity with Korean accented English was 3.4 ($SD = 1.7$) on a scale of 1 (not familiar at all) to 9 (extremely familiar).

Stimulus Description

74 L2 speech samples from 37 Korean speakers (10 male; 27 female) and six L1 speech samples from three English speakers (1 male; 2 female) were used. The mean age of the Korean speakers was 31.5 ($SD = 6.5$). Their length of residence in English speaking countries ranged

from 1 month to 8 years ($M = 2.1$, $SD = 2.1$). The Korean speakers also had a wide range of English proficiency levels, ranging from students in ESL beginner classes to graduate students in the United States who earned close to perfect scores on the internet-based TOEFL⁷. The six L1 speech samples served as reference points to which the listeners could compare the L2 speech. The three English speakers were undergraduate or graduate students at a large university in the United States. The speech samples were responses to the same two questions as in Study 1, one about their major field and the other about their free time activities (Appendix A). For presentation to the raters, 20-second excerpts were taken from approximately the middle of the original recordings (Bosker et al., 2013; Derwing et al., 2007). Each excerpt started and ended at a clause boundary. All the speech samples were normalized in Praat (Boersma & Weenink, 2012) to have a mean intensity of 70dB.

Procedure

The raters heard 80 speech samples in random order over headphones and rated their level of fluency using a nine-point scale with labeled extremes (1 = extremely disfluent, 9 = extremely fluent). The speech excerpts and the scale were presented to raters using Praat (Boersma & Weenink, 2012). The scale appeared on the screen after each sample excerpt had been played; therefore, raters could rate each excerpt only after they heard the whole excerpt. The raters were asked to rate how easily and smoothly speech is delivered, focusing on features of fluency such as speed, pause and repair phenomena, rather than in terms of overall proficiency (see Appendix C for the instructions). Before the actual experiment, each rater completed a practice session to ensure familiarity with the task. In the experiment speech samples were

⁷ The speech samples collected for Study 1 were also included for the stimuli in Study 2.

completely randomized for each rater. The experiment was conducted in a quiet room with a group of at most 4 raters per session. The rating experiment took about 40 minutes and the raters were able to take a short break after rating half of the speech excerpts. After completing the rating experiment, the raters filled out a short questionnaire on their background information, familiarity with Korean accented English, and L2 learning and teaching experience (see Appendix D).

Acoustic Analysis of Speech Excerpts

In order to investigate relationships between fluency ratings and pause phenomena in speech, the L2 speech materials were analyzed acoustically. First, all speech excerpts were transcribed in detail including information regarding silent pauses (≥ 250 ms; De Jong & Bosker, 2013; Kahng, 2012). The length of silent pauses was measured in milliseconds (ms) by listening to each speech excerpt and examining the waveform and spectrogram using Praat (Boersma & Weenink, 2012), and the duration was added to the transcript. Pauses were also categorized, depending on their locations, as either within clauses or between clauses (Foster et al., 2013). Next, the frequency and duration of silent pauses were measured by *Number of silent pauses per minute* and *Mean length of silent pauses*, respectively. The distribution of silent pauses was operationalized by *Silent pause rate within a clause* based on the findings in Study 1 which indicated that, unlike silent pauses at grammatical junctures (e.g., at a clause boundary), the measure not only clearly distinguished between the L1 and L2 speakers but also had a strong negative correlation with L2 oral proficiency. *Silent pause rate within a clause* captures how often a speaker pauses within each clause on average and was computed by dividing the total number of silent pauses occurred within clauses by the number of clauses in each speech excerpt.

Statistical Analysis

To analyze the relative contributions of frequency, length, and distribution of silent pauses to fluency ratings, multiple regression analyses were conducted with fluency ratings as a dependent variable and the three measures on pause phenomena (i.e., *Number of silent pauses per minute*, *Mean length of silent pauses*, and *Silent pause rate within a clause*) as predictor variables. A log transformation was performed on *Mean length of silent pauses* and *Silent pause rate within a clause* so that the data could closely approximate the normal distribution.

Results

The 46 raters evaluated 80 speech excerpts in terms of their level of fluency and the interrater reliability and interrater agreement was high. Cronbach's alpha coefficient was 0.98 and the intraclass correlation coefficient (absolute agreement) was 0.93. I report both Cronbach's alpha coefficient and intraclass correlation coefficient because the former measures internal consistency and reliability of the measure (treating the raters as items; Carr, 2011) and the latter measures the extent to which the individual raters agree with one another in their ratings (Field, 2005). For the intraclass correlation, I used a two-way random model as both the speakers and the raters were random effects (Larsen-Hall, 2010). Table 8 shows the descriptive statistics of *Number of silent pauses per minute*, *Mean length of silent pauses*, and *Silent pause rate within a clause* and fluency ratings of L2 speech excerpts. As expected from a wide range of L2 proficiency, the L2 speakers demonstrated a range of performance in terms of frequency, duration, and distribution of silent pauses in Table 8.

Table 8: Descriptive statistics of pause phenomena and fluency ratings of L2 speech

	<i>N</i>	<i>M</i>	<i>SD</i>	<i>Min.</i>	<i>Max.</i>
<i>Number of silent pauses per minute</i>	74	23.45	6.01	12.52	37.89
<i>Mean length of silent pauses (ms)</i>	74	721	192	390	1360
<i>Silent pause rate within a clause</i>	74	0.93	0.82	0.00	5.50
Fluency ratings	74	5.56	1.53	2.35	8.35

Table 9 shows Pearson correlations between the measures and fluency ratings. The correlation analysis shows that the frequency and length measures are not correlated with each other but the frequency and distribution measures are correlated, which seems natural considering that *Silent pause rate within a clause* is related to the number of pauses within clauses. Correlations between pause phenomena and fluency ratings demonstrated that all three measures are negatively correlated with ratings. In particular, *Silent pause rate within a clause* exhibited the highest correlation with fluency ratings and *Number of silent pauses per minute* had a moderately strong correlation with ratings.

Table 9: Correlations between the measures of pause phenomena and fluency ratings

	SPmin	LngSP	SPRwc	Ratings
<i>Number of silent pauses per minute (SPmin)</i>	1			-.555**
<i>Mean length of silent pauses (LngSP)</i>	-.019	1		-.339**
<i>Silent pause rate within a clause (SPRwc)</i>	.692**	.226	1	-.673**

Note. ** = $p < .01$.

A multiple linear regression analysis was performed in order to investigate to what extent each aspect of pause phenomena can explain the variance of the fluency ratings. First, based on previous findings that pause frequency and duration are related to perceived fluency, the two variables were entered first and the measure of pause distribution was entered last so as to

examine whether pause distribution can explain additional variance of ratings. Table 10 shows the results of the hierarchical multiple regression analysis.

Table 10: Results of a hierarchical multiple regression

Model	Predictors	R^2	R^2 change	F change	df	p
1	Frequency	.291	.291	28.740	1, 70	< .001
2	Frequency + Length	.422	.131	15.615	1, 69	< .001
3	Frequency + Length + Distribution	.519	.097	13.670	1, 68	< .001

Results of the hierarchical multiple regression show that pause frequency explained 29% of the variance of the fluency ratings and when pause length was added, it explained an additional 13% of the variance. Finally, when pause distribution was added, it was able to explain additional 10% of the variance of the fluency ratings, which was significantly more ($p < .001$) than what the model without the measure of pause distribution could explain. The three silent pause measures altogether were able to explain about 52% of the variance of the fluency ratings. In addition, to see the effects of the order in which predictors are entered into the model, a stepwise multiple regression was performed to compare the results based on a mathematical criterion with the results of the hierarchical multiple regression.

Table 11 shows that with the stepwise multiple regression, as pause distribution had the highest correlation with fluency ratings (Table 9), it was entered into the model first and was able to explain over 45% of the variance of the fluency ratings by itself. Next, pause length was entered and it explained additional 4% of the variance; however, pause frequency was not included in the model as it did not explain significantly additional variance of the fluency ratings.

Table 11: Results of a stepwise multiple regression

Model	Predictors	R^2	R^2 change	F change	df	p
1	Distribution	.452	.452	57.662	1, 70	< .001
2	Distribution + Length	.493	.041	5.751	1, 69	.021

Discussion

Experiment 1 examined the relative contributions of frequency, length and distribution of silent pauses to perceived fluency. The first research question of Experiment 1 was which acoustic measures of pause phenomena (frequency, duration and/or distribution of silent pauses) are significantly related to fluency ratings. The results showed that fluency ratings were significantly correlated with all three measures—frequency ($r = -.555$), duration ($r = -.339$), and distribution ($r = -.673$) of silent pauses. It is especially noteworthy that pause distribution exhibited the strongest correlation with fluency ratings among the three pause variables. The second research question of Experiment 1 was whether pause distribution explains significantly additional variance of fluency ratings which is not explained by frequency and duration of silent pauses. The hierarchical multiple regression analysis showed that the regression model with pause frequency and length explained 42% of the variance of fluency ratings and when pause distribution was added to the model it was able to explain about 10% of additional variance of fluency ratings.

Although it has been suggested that pause phenomena play an important role in the perception of fluency, previous studies have been equivocal on the relationship between perceived fluency, and frequency and length of silent pauses. Pause frequency was correlated with fluency ratings in Rossiter (2009) but not in Kormos & Dénes (2004); whereas pause

duration was correlated with fluency ratings in Kormos & Dénes (2004) but not in Cucchiari et al. (2002). Moreover, neither the relationship between perceived fluency and pause distribution, nor the relative contributions of frequency, duration, and distribution of pauses had been investigated.

Experiment 1 precisely examined these gaps in the literature and predicted that not only pause frequency and length but also pause distribution would be significantly correlated with fluency ratings, based on the L1 research findings that pause location influences speech perception (e.g., Arons, 1993; Griffiths, 1991; Reich, 1980, Seidl & Cristià, 2008; Sugito, 1990). Silent pauses are one of the acoustic cues to clausal units (Seidl & Cristià, 2008) and sentences containing silent pauses between clauses were processed faster and recalled better than sentences containing silent pauses within clauses (Reich, 1980). In Experiment 1 pause distribution was operationalized by *Silent pause rate within clauses* (i.e., number of within-clause pauses per clause), which demonstrated a striking difference between L1 and L2 speech in Study 1. The results followed the prediction by demonstrating correlations between fluency ratings and pause frequency, length, and distribution. Pause distribution was also able to explain 10% of additional variance that was not explained by pause frequency and length. Given that pause distribution and pause frequency had a strong correlation ($r = -.673$), the additional 10% of explanatory power that pause distribution had suggests its crucial role in perceived fluency. In fact, pause distribution exhibited the strongest correlation with fluency ratings and was able to explain 45% of the variance of fluency ratings. Moreover, when pause distribution was entered into the regression model first, pause frequency was not able to explain additional variance of fluency ratings.

Experiment 2

Experiment 2 tested a causal relationship between pause location and perceived fluency through speech manipulations. The results of Study 1 have shown that one of the major differences between L1 and L2 speakers' speech lies in the frequency of pauses within clauses. Although frequency and length of silent pauses have been reported to be correlated with fluency ratings (e.g., Kormos & Dénes, 2004; Rossiter, 2009), effects of pause location on perceived fluency have not been investigated. Therefore, the experiment specifically aims to answer whether pauses within clauses decrease fluency ratings compared to pauses between clauses.

Furthermore, although L1 speakers also produce disfluencies (e.g., pauses and repairs), L1 speakers tend to be perceived as fluent by default (Davies, 2003; Riggenbach, 1991), and studies investigating the relationship between utterance fluency and perceived fluency of L1 speakers are rare. Bosker (2013) recently compared the way raters evaluate fluency of L1 and L2 speech. He manipulated L1 and L2 speech in terms of pauses, by constructing no pause, short pause, and long pause conditions, and speed, by speeding up L2 speech and slowing down L1 speech. The results showed that the ratings of manipulated L1 and L2 speech were affected in a similar fashion, suggesting that listeners evaluate fluency characteristics of L1 and L2 speakers in a similar way. Bosker (2013) also has methodological implications. Many previous studies used correlational analyses to explore the relationship between utterance fluency and perceived fluency (e.g., Bosker et al., 2013; Cucchiarini et al., 2002; Derwing et al., 2004; Kormos & Dénes, 2004; Rossiter, 2009). However, Bosker (2013) points out that the correlational approach would be unsuitable to compare the perception of L1 and L2 speech because they differ in many respects. Hypothetically, if pause frequency is found to be more strongly correlated with ratings of L2 speech than with ratings of L1 speech, it could be due to the fact that L2 speech had more

pauses as compared to L1 speech, and not due to a difference in relative weight of pausing.

Therefore, he used phonetic manipulations to ascertain that the effects on fluency ratings could be directly attributed to the fluency characteristics they manipulated, and that he could compare how the same modification in L1 and L2 speech affects perceived fluency.

Building upon the previous studies, Experiment 2 aims to fill gaps and extend the body of research on perceived fluency. The role of pause location in L1 and L2 perceived fluency has not yet been investigated. Therefore, Experiment 2 examined a causal relationship between pause location and perceived fluency by constructing three conditions—*No Pause*, *Pauses Between Clauses*, and *Pauses Within Clauses* conditions—and compared fluency ratings of L1 and L2 speech in the three conditions. These conditions were created through phonetic manipulations of L1 and L2 speech in order to directly test for a causal effect of pause location on perceived fluency of L1 and L2 speech. The research questions are as follows.

1. Is there a difference in fluency ratings of L1 speech when the speech has 1) no pause, 2) pauses between clauses, and 3) pauses within clauses?
2. Is there a difference in fluency ratings of L2 speech when the speech has 1) no pause, 2) pauses between clauses, and 3) pauses within clauses?

In Bosker (2013), both L1 and L2 speech in the no pause condition were rated as more fluent than the short and long pause conditions. Therefore, in Experiment 2 of the current study, both L1 and L2 speech in the *No Pause* condition are also predicted to be rated as more fluent than the *Pauses Between Clauses* and *Pauses Within Clauses* conditions. Regarding the difference in ratings between the *Pauses Between Clauses* and *Pauses Within Clauses* conditions,

as pauses are one of the acoustic cues to clausal units (Seidl & Cristià, 2008), raters are likely to prefer pauses between clauses to pauses within clauses. Perception of L1 and L2 speech might be influenced by pause location in the same way as it was by speed, and pause frequency and length in Bosker (2013). However, based on previous research on L1 pause phenomena, there is still a possibility that pause location affects perception of L1 and L2 speech to a different degree. L1 research has shown that, only under conditions with cognitive complexity, do silent pauses between clauses have apparent beneficial effects on speech processing, whereas silent pauses within clauses interfere with speech perception processing and recall (Arons, 1993; Bower & Springston, 1970; Griffiths, 1991; Reich, 1980; Sugito, 1990). Moreover, a number of studies on pause detection in L1 speech show that listeners are not good at detecting disfluencies such as pauses and repairs (Bailey & Ferreira, 2003). For instance, in transcription tasks, listeners tend to misplace pauses within clauses to between clauses (e.g., Duez, 1985; Martin & Strange, 1968) and Duez (1985) concluded that listeners tend not to hear pauses which are not expected, such as within-constituent pauses. Given that it is possible that raters may consider perception of L2 speech cognitively more demanding than that of L1 speech, and that raters may not expect to hear pauses within clauses in L1 speech but they may do in L2 speech, in Experiment 2 it is predicted that fluency ratings of L2 speech can be influenced by pause location more than those in L1 speech.

Method

Raters

Ninety-two native English speakers (20 male; 72 female) participated in the study as raters. They were undergraduate students at a large university in the United States and the mean

age was 21 ($SD = 2.0$) and reported to have normal hearing. Their mean familiarity with Korean accented English was 3.7 ($SD = 2.0$) on a scale of 1 (not familiar at all) to 9 (extremely familiar).

Stimulus Description

Twenty-four L1 and 24 L2 spontaneous speech samples recorded by 12 English speakers and 12 Korean learners of English were used, which had been collected for Study1 and Experiment 1. The speech samples were responses to two questions, one about the speaker's major field and the other about their free time activities (Appendix A). The samples were selected so that the L1 and L2 speech samples were comparable in terms of speed fluency; there was no significant group difference in *Mean syllable duration* ($M_{L1} = 246$, $SD_{L1} = 23$; $M_{L2} = 263$, $SD_{L2} = 22$). Fragments of approximately 20 seconds were excerpted from the middle of the original recordings (Bosker et al., 2013; Derwing et al., 2007). Each excerpt started and ended at a clause boundary.

Three conditions—‘*No Pause*,’ ‘*Pauses Between Clauses*,’ and ‘*Pauses Within Clauses*’—were created to test whether pauses within clauses lower fluency ratings compared to pauses between clauses or no pause. To test for a causal relationship between pause location and perceived fluency, the speech samples in the two conditions with pauses (i.e., *Pauses Between Clauses* and *Pauses Within Clauses*) should be different only in terms of pause location but should have the same number of pauses with the same length. Therefore, first, to create the *No Pause* condition, all the silent pauses in the speech samples were shortened to the length of around 150 milliseconds (Bosker, 2013). Next, stimuli for the *Pauses Between Clauses* and *Pauses Within Clauses* conditions were constructed by adding the same number of pauses with

the same length either between clauses or within clauses depending on the condition, to the speech samples in the *No Pause* condition. After examining all the speech samples, it was decided to add five pauses to them. Five was the optimal number of pauses in that all the speech samples could have five pauses within and between clauses naturally, without interrupting coarticulation. The length of pauses added was around 600 milliseconds, which was about the average length of English speakers' silent pauses in Study 1.

A clause was required to consist minimally of a finite or non-finite verb with at least one other clause element such as a subject, object, or complement (see Foster et al., 2000, pp. 365-368). Examples of pauses between clauses are: *I performed in several plays [pause] I believe [pause] I have some talent in acting*. Examples of pauses within a clause are: *learn new [pause] things; so [pause] hard, to my [pause] place* (see Appendix E and F for more examples).

The speech samples were normalized in Praat (Boersma & Weenink, 2012) to have a mean intensity of 70dB. In addition, a small subtle white noise was added (33dB) to the speech samples using the RandomGauss function in Praat ($M = 0$, $SD = 0.001$). This was done in order to normalize the background noise throughout and across the speech samples in an attempt to mask any possible trace of pause manipulations. The level of noise was very low and sounded like part of the original recordings; therefore, none of the raters noticed that a noise had been added to the speech samples. All the manipulated speech samples were evaluated for naturalness by two native English speakers and two advanced learners of English and corrections were made, if necessary (e.g., changing pause locations). All the locations where a pause was added originally had a silence; therefore, none of the added pauses interrupted coarticulation. The stimuli were arranged according to a Latin Square design, in which raters were presented with each item in only one condition, with three groups of raters for counterbalancing. A Latin Square

design was used because when raters hear the same speech excerpts more than once, the familiarity with the content of the speech excerpts is likely to affect their ratings. Table 12 demonstrates how speech samples were organized according to a 3 x 3 Latin Square design. 24 speakers were randomly assigned to one of the three speaker groups (i.e., S1, S2, S3) and each speaker group consisted of 4 L1 speakers and 4 L2 speakers. Ninety-two raters were also randomly assigned to one of the three rater groups (i.e., R1, R2, R3). For instance, raters in R1 heard speech samples of S1 in the *No Pause* condition, speech samples of S2 in the *Pauses Between Clauses* condition, and speech samples of S3 in the *Pauses Within Clauses* condition. By doing so, raters listened to each speech sample in only one condition.

Table 12: A schematic representation of the 3 x 3 Latin Square design. No, B, and W represent the *No Pause*, *Pauses Between Clauses*, and *Pauses Within Clauses* conditions, respectively.

		Speakers		
		S1	S2	S3
Raters	R1	No	B	W
	R2	W	No	B
	R3	B	W	No

Procedure

As detailed above, 92 raters were randomly assigned to one of the three rater groups for counterbalancing. Each rater heard 48 manipulated speech samples produced by 24 speakers in random order over headphones and rated the level of fluency of the speaker using a nine-point scale with labeled extremes (1 = extremely disfluent, 9 = extremely fluent). As in Experiment 1, the speech excerpts and the scale were presented to raters using Praat (Boersma & Weenink, 2012). The scale appeared on the screen after each sample excerpt had been played; therefore,

raters could rate each excerpt only after they heard it completely. The raters were asked to rate how easily and smoothly speech is delivered, focusing on features of fluency such as speed, pause and repair phenomena, rather than in terms of overall proficiency (see Appendix C for the instructions). Before the actual experiment, there were three practice items so that raters could familiarize themselves with the procedure. In the experiment speech samples were randomized for each rater. The procedure was conducted in a quiet room with a group of at most 4 raters per session. The rating experiment took about 35 minutes and the raters were able to take a short break after rating half of the speech excerpts. After they finished rating, they filled out a short questionnaire on their background information, familiarity with Korean accented English and L2 learning and teaching experience (Appendix D). Lastly, they were also asked whether they had noticed anything particular or interesting about the speech excerpts and none of them mentioned that the speech samples sounded unnatural or manipulated.

Analysis

The interrater agreement within the three rater groups was high (Cronbach's alpha coefficients: 0.94, 0.93, 0.95; intraclass correlation coefficients in terms of absolute agreement: 0.89, 0.84, 0.91⁸). In order to test whether the three pause conditions affected fluency ratings of L1 and L2 speakers' speech, mixed effects ANOVAs were performed with fluency ratings of L1 and L2 speakers' speech as dependent variables using SPSS Statistics 17.0 (SPSS Inc., 2008). Mixed effects ANOVAs were performed in order to test effects of fixed variable (i.e., Condition)

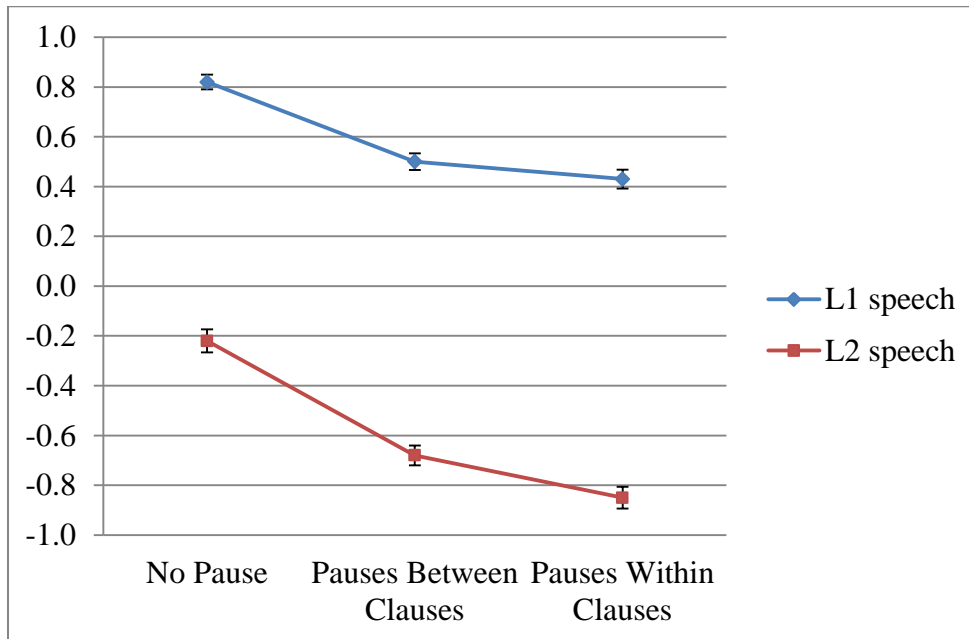
⁸ Intraclass correlation coefficients (ICC) seem to be a bit lower than Cronbach's alpha coefficients as the ICCs measured the extent of absolute agreement across raters. The intraclass correlation can be considered to be a conservative estimate of interrater reliability (Stemler & Tsai, 2007).

more accurately while taking into account effects of random variables such as Speaker group and Rater group. The mean ratings across raters were not the same, and rating responses were relative rather than absolute, for instance, one rater's 7 on the 9-point scale is not likely to be the same as other raters' response of 7 on the 9-point scale. Therefore, fluency ratings were standardized by calculating z-scores using each rater's mean and standard deviation for a more accurate analysis by addressing individual differences in ratings. The transformed data closely approximated normal distributions.

Results

Figure 4 illustrates the means and standard errors of fluency ratings of L1 and L2 speech in the three conditions. First, the figure shows that the L1 speech excerpts were rated higher than the L2 speech excerpts. It also shows that for both L1 and L2 speech, ratings of the *Pauses Between Clauses* and *Pauses Within Clauses* conditions are lower than ratings of the *No Pause* condition. Ratings of the *Pauses Within Clauses* condition seem lower than those of the *Pauses Between Clauses* condition for both L1 and L2 speech; however, the difference between the two conditions seems larger for L2 speech. In order to examine statistical differences between the three conditions for L1 and L2 speech, mixed effects ANOVAs were conducted with fluency ratings of L1 and L2 speakers' speech as dependent variables.

Figure 4: Mean and standard error z-scores of fluency ratings of L1 and L2 speech



The first mixed effects ANOVA was run with ratings of L1 speech excerpts as a dependent variable, Condition as a fixed variable, and Speaker group, Rater group, and Raters within rater groups as random variables. The results showed that there was a main effect of Condition ($F(2, 1008) = 42.790, p < .001, \eta_p^2 = .078$), Speaker group ($F(2, 1008) = 20.746, p < .001, \eta_p^2 = .040$), Raters within rater groups ($F(89, 1008) = 1.441, p = .006, \eta_p^2 = .113$), and Rater group ($F(2, 89) = 3.178, p = .046, \eta_p^2 = .067$). In order to compare ratings between the three conditions, post hoc tests were performed using Tukey HSD. The results showed that the L1 speech excerpts in the *Pauses Between Clauses* condition ($p < .001$) and in the *Pauses Within Clauses* condition ($p < .001$) were rated significantly lower than the samples in the *No Pause* condition. However, there was no significant difference in ratings between the *Pauses Between*

Clauses and *Pauses Within Clauses* condition ($p = .247$). Therefore, L1 speech samples were rated lower when they had pauses either within or between clauses than when they had no pauses; however, L1 speech samples containing pauses within clauses and between clauses were not significantly different in fluency ratings.

Next, another mixed effects ANOVA was run with ratings of L2 speech excerpts as a dependent variable, Condition as a fixed variable, and Speaker group, Rater group, and Raters within rater groups as random variables. There was a main effect of Condition ($F(2, 1008) = 56.728, p < .001$), Speaker group ($F(2, 1008) = 48.870, p < .001$), and Rater group ($F(2, 89) = 3.178, p = .046$); however, no significant effect of Raters within rater groups ($F(89, 1008) = 0.801, p > .05$). Therefore, in order to build a model that explains the data better, Condition, Speaker group, and Rater group were included and Raters within rater groups was excluded. The new model showed that there was a main effect of Condition ($F(2, 1097) = 57.660, p < .001$) and Speaker group ($F(2, 1097) = 49.673, p < .001$); however, no significant effect of Rater group ($F(2, 1097) = 2.587, p > .05$). Finally, the model which further excluded Rater group showed that there was a main effect of Condition ($F(2, 1099) = 57.494, p < .001, \eta_p^2 = .095$) and Speaker group ($F(2, 1099) = 49.530, p < .001, \eta_p^2 = .083$). To examine significant differences between the three conditions, Tukey HSD post hoc tests were performed. The results showed that as in L1 speech ratings, L2 speech samples in the *Pauses Between Clauses* ($p < .001$) and *Pauses Within Clauses* conditions ($p < .001$) were rated significantly lower than L2 speech samples in the *No Pause* condition. However, unlike L1 speech samples, L2 speech samples in the *Pauses Within Clauses* condition were rated significantly lower than those in the *Pauses Between Clauses* condition ($p = .011$).

In addition, in order to further explore possible acoustic factors of this differential effect of pause location on perceived fluency of L1 and L2 speech, the speech excerpts were analyzed and compared to investigate whether there were differences between them in terms of fluency features. The research design of Experiment 2 ensured that the L1 and L2 speech excerpts were comparable in terms of the two major oral correlates of perceived fluency—speed and silent pauses. On the other hand, other minor oral correlates of perceived fluency such as frequency of filled pauses and repairs have not been matched between the L1 and L2 speech samples. Therefore, the number of filled pauses and the number of repairs in the L1 and L2 speech excerpts were calculated and compared. The results showed that the L1 and L2 speech were comparable in the number of filled pauses ($M_{L1} = 2.04$, $M_{L2} = 1.96$, $SD_{L1} = 1.18$, $SD_{L2} = 1.67$, $t(22) = 0.141$, $p = .889$) but the L2 speech had more repairs than the L1 speech ($M_{L1} = 0.17$, $M_{L2} = 0.67$, $SD_{L1} = 0.33$, $SD_{L2} = 0.69$, $t(15.727) = -2.283$, $p = .037$, $r = .50$).

Discussion

Experiment 2 examined whether pause location influences perceived fluency of L1 and L2 speech. In order to test a causal effect of pause location on fluency ratings of L1 and L2 speech, three conditions were constructed—*No Pause*, *Pauses Between Clauses*, and *Pauses Within Clauses* conditions. The conditions were created through phonetic manipulations. The baseline *No Pause* condition was created by shortening all the silent pauses in the speech samples to the length of around 150 milliseconds (Bosker, 2013). To examine effects of pause location on perceived fluency directly, the speech samples in the *Pauses Between Clauses* and *Pauses Within Clauses* conditions were prepared by adding the same number of pauses with the

same length, either within clauses or at clause boundaries depending on the condition, to the speech samples in the *No Pause* condition. The research question was whether there is a difference in fluency ratings when the L1 and L2 speech have 1) no pause, 2) pauses between clauses, and 3) pauses within clauses.

Based on the findings in Bosker (2013) that both L1 and L2 speech in the no pause condition were rated as more fluent than the short and long pause conditions, in Experiment 2, both L1 and L2 speech in the baseline *No Pause* condition were also predicted to be rated as more fluent than the conditions with pauses—*Pauses Between Clauses* and *Pauses Within Clauses* conditions. The results followed the prediction and showed that both L1 and L2 speech in the *No Pause* condition were rated to be more fluent than those in the *Pauses Between Clauses* and *Pauses Within Clauses* conditions.

Regarding the main focus of Experiment 2, the effect of pause location on fluency ratings, raters were predicted to prefer pauses between clauses to pauses within clauses in general, as pauses are one of the acoustic cues to clausal units (Seidl & Cristià, 2008). The perception of L1 and L2 speech might be influenced by pause location to a similar degree, as it was by speed, and pause frequency and length in Bosker (2013). However, it was predicted that pause location would affect the perception of L2 speech more than that of L1 speech, based on previous research on L1 pause phenomena (e.g., Arons, 1993; Bower & Springston, 1970; Griffiths, 1991; Reich, 1980; Sugito, 1990), which suggested that effects of pause location on speech perception are apparent under cognitively demanding conditions, and listeners tend not to hear pauses which are not expected, such as within-constituent pauses (Duez, 1985).

The results of the effects of pause location on the fluency ratings also followed the predictions. Although both L1 and L2 speech had lower fluency ratings in the *Pauses Within*

Clauses condition than in the *Pauses Between Clauses* condition, only the ratings of L2 speech were significantly affected by pause location. The difference in the ratings of L1 speech in the two conditions did not reach significance. The current findings suggest that overall, listeners are sensitive to pause location. The results are compatible with the fact that a silent pause is an acoustic cue to clausal units (Seidl & Cristià, 2008). The L1 infant studies also have shown that 6-month old infants prefer sentences containing pauses between clauses to those containing pauses within clauses (Hollich & Houston, 2007). More interestingly, the results showed that perceived fluency of L2 speech was influenced by pause location more than that of L1 speech. This differential effect of pause location on L1 and L2 perceived fluency is consistent with the L1 literature which showed that silent pauses have beneficial effects on listeners only under conditions of cognitive complexity in auditory speech processing and they did not demonstrate apparent beneficial effects when the speech or tasks were easy enough (Aaronson, 1968; Reich, 1980). It is possible that raters considered perception of L2 speech more cognitively demanding than that of L1 speech. The finding is also in agreement with L1 pause detection studies (e.g., Duez, 1985; Martin & Strange, 1968) and Duez's (1985) claim that listeners tend not to hear pauses which are not expected, such as within-constituent pauses. Raters may not have expected to hear pauses within clauses in L1 speech; however, they may have expected to hear pauses within clauses in L2 speech and thus, may have been more ready and sensitive to detect them in L2 speech than in L1 speech.

In addition, in an attempt to further explore possible acoustic factors of this differential effect of pause location on perceived fluency of L1 and L2 speech, the excerpts were analyzed and compared to investigate whether there were differences between them in terms of fluency characteristics. The research design of Experiment 2 ensured that the L1 and L2 speech excerpts

were comparable in terms of the two major oral correlates of perceived fluency—speed and silent pauses. As discussed in the method section, the L1 and L2 speech samples were selected so that the two groups were comparable in terms of speed (i.e., *Mean syllable duration*). Through the phonetic manipulations, both L1 and L2 speech excerpts had no silent pauses ($\geq 250\text{ms}$) in the *No Pause* condition, and they had the same number of pauses with the same length in both *Pauses Between Clauses* and *Pauses Within Clauses* conditions. On the other hand, other minor oral correlates of perceived fluency such as frequency of filled pauses and repairs have not been matched between the L1 and L2 speech samples. Acoustic analyses showed that the L1 and L2 speech were comparable in the number of filled pauses but the L2 speech had more repairs than the L1 speech. The findings suggest that pause location affected fluency ratings more when speech had more repairs. It is possible that repairs in speech have made speech perception more cognitively demanding. Another possibility is that repairs in speech have led listeners to expect more pauses within clauses and to be sensitive and ready to detect them, affecting their ratings. Further research is needed to confirm whether and how repairs in speech interact with pause location and affect perceived fluency. In addition, it should be noted that there were differences between the L1 and L2 speech other than fluency characteristics such as accent, linguistic accuracy and complexity, and whether these differences can interact with pause location and affect perceived fluency also requires further investigation.

General Discussion: Study 2

In order to find the speech features that influence L2 perceived fluency, a number of studies have investigated the relationship between utterance fluency and perceived fluency (e.g., Bosker et al., 2013; Cucchiarini, Strik, & Boves, 2000, 2002; Derwing, Rossiter, Munro, &

Thomson, 2004; Fulcher, 1996; Kormos & Dénes, 2004; Rossiter, 2009) and suggested importance of silent pauses on perceived fluency; however, both the role of pause location in perceived fluency, and the relative contributions of the frequency, length, and distribution of silent pauses to perceived fluency have rarely been examined. The current study aimed to fill these gaps and extend the body of research on L2 perceived fluency using two experiments.

Experiment 1 investigated the relative contributions of the frequency, length, and distribution of silent pauses on L2 perceived fluency and showed that pause distribution, in particular exhibited the strongest correlation with fluency ratings and explained 45% of the variance of fluency ratings. Pause distribution was also able to explain 10% of additional variance that was not explained by pause frequency and length, suggesting its crucial role in perceived fluency. Experiment 2 tested whether pause location affected perceived fluency of L1 and L2 speech by comparing fluency ratings of L1 and L2 speech in the three conditions—*No Pause*, *Pauses Between Clauses*, and *Pauses Within Clauses*. The results showed that both L1 and L2 speech were rated higher when they had no pause than when they had pauses. More importantly, L1 and L2 speech in the *Pauses Within Clauses* condition were rated lower than those in the *Pauses Between Clauses* condition; however, only the difference in ratings of L2 speech reached significance.

Findings of both Experiment 1 and Experiment 2 suggest a significant role of pause location on L2 perceived fluency. They are consistent with L1 research on pause phenomena which suggests that pause location affects speech perception. Silent pauses are one of the acoustic cues to clausal units along with pitch and vowel duration (Seidl & Cristià, 2008). Therefore, silent pauses at grammatical boundaries help listener comprehension by indicating the boundaries of speech to be analyzed, and by providing cognitive processing time (e.g., Arons,

1993; Griffiths, 1991; Reich, 1980, Sugito, 1990), whereas pauses within clauses can be disrupting. It has also been reported that silent pauses between clauses have beneficial effects on listeners only under conditions of cognitive complexity in auditory speech processing and they did not demonstrate apparent beneficial effects when the speech or tasks were easy enough (Aaronson, 1968; Reich, 1980). It is possible that the raters in Experiment 2 considered perception of L2 speech more cognitively demanding than that of L1 speech. The finding is also compatible with L1 pause detection studies (e.g., Duez, 1985; Martin & Strange, 1968) and Duez's (1985) claim that listeners tend not to hear pauses which are not expected, such as within-constituent pauses. The raters in Experiment 2 may not have expected to hear pauses within clauses in L1 speech; however, they may have expected to hear pauses within clauses in L2 speech and thus, may have been more ready and sensitive to detect them in L2 speech than in L1 speech. The findings of Study 2 can be viewed as an initial attempt to fill the gaps in the literature on the relationship between L2 perceived fluency and pause location and further research is needed in particular on which aspects of L2 speech make perceived fluency more susceptible to pause location than those of L1 speech.

Study 2 also has methodological implications. Experiment 2 used phonetic manipulations to test a causal relationship between pause location and perceived fluency. The correlational approach would be unsuitable to compare the perception of L1 and L2 speech because they differ in many respects. Phonetic manipulations ensured that effects in fluency ratings could be directly attributed to fluency characteristics manipulated in both L1 and L2 speech.

CHAPTER 4: CONCLUSION

Fluency is one of the most noticeable differences between native and nonnative speech and constitutes a critical component of second language (L2) proficiency; however, the concept has not been well understood by researchers. In order to deepen understanding of the multidimensional construct of fluency, the current dissertation took a novel approach and investigated the production and perception of second language fluency from all three aspects of fluency—utterance, cognitive, and perceived fluency.

Study 1 investigated utterance fluency and cognitive fluency of English speakers and Korean learners of English by comparing temporal measures and stimulated recall responses. The L1 and L2 speakers were different in speed, length of run, repairs, and silent pauses. In particular, a striking group difference in *Silent pause rate within a clause* is consistent with the claim that pauses within clauses reflect processing difficulties in speech production. Stimulated recall responses showed that lower proficiency learners remembered more issues regarding L2 declarative knowledge on grammar and vocabulary than higher proficiency learners, which was compatible with the declarative/procedural model and studies on automaticity.

Study 2 examined the relationship between utterance fluency and perceived fluency using two experiments. Experiment 1 investigated the relative contributions of frequency, length, and distribution of pauses to perceived fluency of L2 speech. Experiment 2 tested causal effects of pause location on perceived fluency of L1 and L2 speech. Findings of both Experiment 1 and Experiment 2 suggest a significant role of pause location on L2 perceived fluency. In Experiment 1, pause distribution demonstrated the strongest correlation with fluency ratings and in Experiment 2, perceived fluency of L2 speech was influenced by pause location more than that

of L1 speech. The findings are in agreement with L1 literature on pause phenomena that silent pauses are one of the acoustic cues to clausal units and silent pauses between clauses can facilitate speech perception and recall, whereas pauses within clauses can interfere with them in cognitively demanding contexts.

The present study has theoretical, methodological, and practical implications in the fields of L2 acquisition research, education, and testing. In terms of theoretical contributions, the study investigated three notions of fluency—utterance, cognitive, and perceived fluency—and examined their relationships in a comprehensive and systematic way within a theoretical framework (Segalowitz, 2010). It also critically identified gaps and issues in previous studies on L2 fluency. An almost exclusive focus on L2 speech in most studies provided little evidence to show how L1 and L2 utterance fluency are different. The relative contributions of frequency, duration, and distribution of silent pauses to perceived fluency have not yet been understood. In particular, effects of pause location on the perception of fluency have rarely been researched. The present study precisely addressed these gaps and aimed to extend the body of research on L2 fluency. Moreover, it took an interdisciplinary approach to capture the multidimensionality of fluency by integrating findings from different fields such as second language acquisition, psycholinguistics, cognitive science, speech science, and pausology.

The two studies also have implications for research methodology. Study 1 discussed strengths and weaknesses of different measures used in the previous studies on utterance fluency, and further proposed measures which could depict pause distribution more accurately. Study 1 also utilized qualitative analysis in exploring cognitive fluency to provide additional insight to the field where quantitative analysis is dominant. Furthermore, in examining effects of temporal features on perceived fluency, Study 2 tried to overcome the limitation of correlation analysis

used in the literature and tested a causal relationship between utterance and perceived fluency using phonetic manipulations.

Results of the studies also have potential implications for L2 education and assessment. Finding reliable oral correlates of fluency can help to improve learners' oral fluency and to develop a more valid assessment tool to measure oral fluency and proficiency in L2 speech. One of the most novel and important findings of the current dissertation is the close relationship between L2 fluency and pauses within clauses. L1 and L2 speech exhibited a striking difference in the frequency of pauses within clauses, which is considered to reflect difficulties in speech production processing such as lexical retrieval. Pauses within clauses also had a crucial impact on perceived fluency of L2 speech. Based on the findings, in classroom one of the ways teachers can help L2 learners to enhance L2 fluency is to provide ample opportunities to practice collocations and formulaic language, which can enable learners to produce longer fluent runs and decrease pauses within clauses in their speech. In terms of L2 assessment, including the measure which addresses the frequency of silent pauses within clauses in automatic fluency assessment can evaluate L2 learners' oral fluency more accurately.

APPENDICES

Appendix A: Questions for spontaneous speech

1. What is your major? What is it about? Do you like it? And why or why not?
2. What do you like to do in your free time?

Appendix B: English language learning background questionnaire in Study 1

1. Age: _____
2. Gender: Male Female
3. Mother tongue (First language): _____
4. Other languages spoken at home as a child: _____
5. Age at first exposure to English
 - a. through instruction: _____
 - b. through immersion-type environment (living in an English-speaking country): _____
6. Years of total instruction (i.e., language courses and content-based coursework) in English up to present day: _____
7. Years of total immersion/exposure to English (i.e., speaking it at home with native speaker(s) and/or living in an English-speaking country): _____
8. First two years of English learning
 - a. How many hours did you have oral English input from native English speakers (i.e., listening to audio materials, speaking with native English speakers) per week?

 - b. How many hours did you have oral interaction with native English speakers per week?

 - c. What was the ratio of oral to written input (e.g., 10:90, 15:85, 75:25)?

9. English proficiency (Please recall as best as you can.)
Test: _____ Total score: _____ Speaking score: _____

Appendix C: Instructions for the experiment

Your task is to listen to native and nonnative speech samples and rate them in terms of their fluency using a 9-point scale.

1: extremely disfluent

9: extremely fluent

In this study fluency refers to **how easily and smoothly speech is delivered**, not overall proficiency. Please make your judgments based on factors such as

- speech rate
- silent and filled pauses (e.g., um, uh)
- hesitations and/or corrections
- overall flow of speech
- NOT grammar or vocabulary

Following are the two questions the speakers answered.

1. What is your major? What is it about? Do you like it? Why or why not?
2. What do you like to do in your free time?

Each stimulus is about 20 second long and was excerpted from approximately the middle of the original recordings.

10. Have you taught or tutored nonnative English speakers? If so, briefly describe your teaching experience (i.e., taught what, to whom, for how long etc.).

11. Which factors do you think particularly influenced your fluency rating? (e.g., speed, silent/filled pauses, repetitions, corrections, accent, grammar, vocabulary etc.) Do you have any other comments about the experiment?

Appendix E: An example of addition of pauses to a speech sample

In my free time which is [PWC] very limited now that I'm a graduate student [PBC]

I [PWC] like to do yoga [PBC] ahm [PBC] or go running or biking [PBC]

Um I also really like to [PWC] cook [PBC]

Ahm which I do [PWC] almost every day but not [PWC] too much

Note. [PWC] represents a pause within a clause and [PBC] represents a pause between clauses.

Appendix F: Example waveforms of the speech manipulations

Figure 5: Speech in the *No Pause* condition

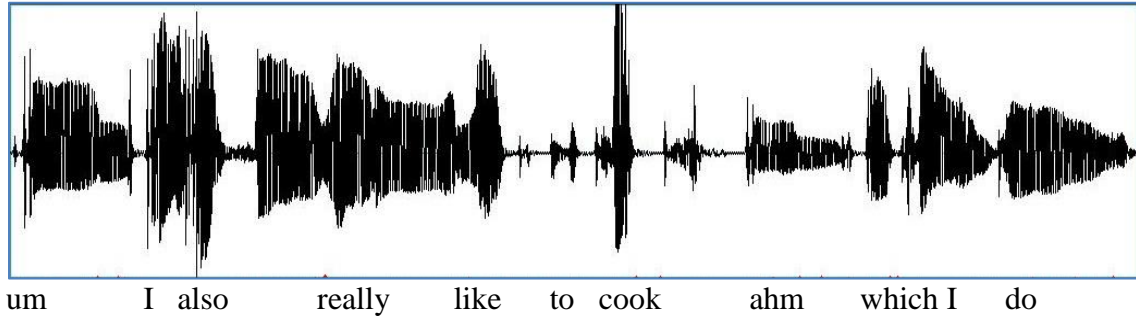


Figure 6: Speech in the *Pauses Between Clauses* condition

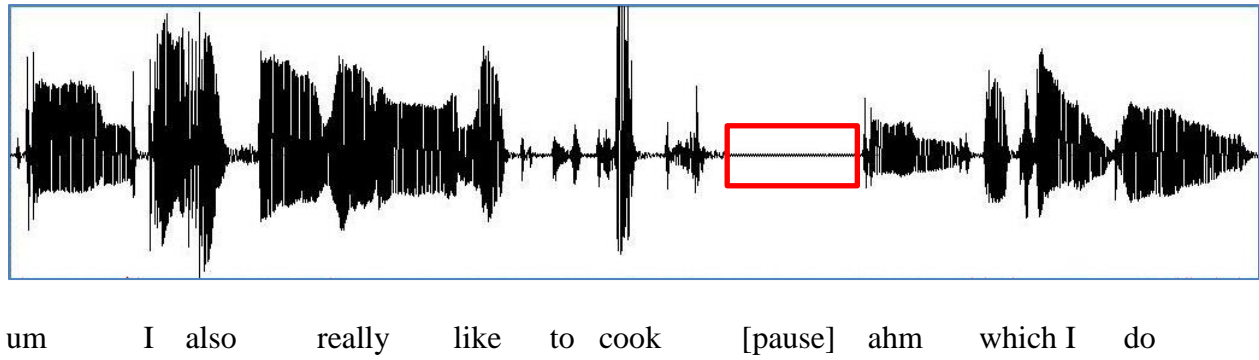
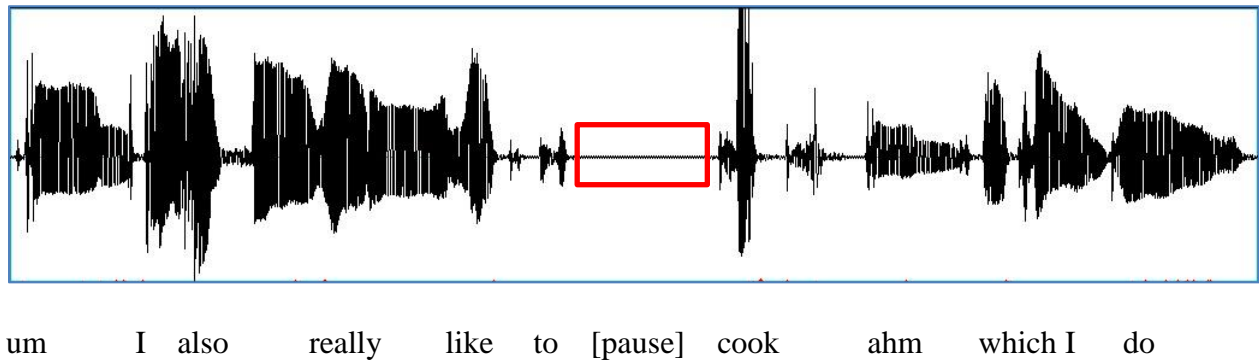


Figure 7: Speech in the *Pauses Within Clauses* condition



REFERENCES

REFERENCES

- Aaronson, D. (1968). Temporal course of perception in an immediate recall task. *Journal of Experimental Psychology*, 76, 129-140.
- Anderson, J. R. (1983). *The architecture of cognition*. Mahwah, NJ: Erlbaum.
- Arons, B. (1993). SpeechSkimmer: Interactively skimming recorded speech. *Proceedings of the 6th Annual ACM Symposium on User Interface Software and Technology, USA*, 6, 187-196.
- Beattie, G. (1977). The dynamics of interruption and the filled pause. *The British Journal of Social and Clinical Psychology*, 16, 283-284.
- Beattie, G. (1980). The role of language production processes in the organization of behavior in face-to-face interaction. In B. Butterworth (Ed.), *Language production: Vol. 1. Speech and talk*, (pp. 69-109). London: Academic Press.
- Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, H. (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. *Language Teaching Research*, 10, 245–261.
- Boersma, P., & Weenink, D. (2012). PRAAT. Retrieved from <http://www.praat.org>
- Boomer, D. S. (1965). Hesitation and grammatical encoding. *Language and Speech*, 8, 148-158.
- Bosker, H. R. (2013, May). *Native and non-native fluency*. Paper presented at the New Sounds 2013 Conference, Montreal.
- Bosker, H. R., Pinger, A., Quené, H., Sanders, T., & de Jong N. H. (2013). What makes speech sound fluent? The contributions of pauses, speed and repairs, *Language Testing*, 30, 159-175.
- Bower, G. H., & Springston, F. (1970). Pauses as recoding points in letter series. *Journal of Experimental Psychology*, 83, 421-430.
- Bybee, J. (2002). Phonological evidence of exemplar storage of multiword sequences. *Studies in Second Language Acquisition*, 24, 215–221.
- Carr, N. T. (2011). *Designing and analyzing language tests*. Oxford: Oxford University Press.
- Cenoz, J. (1998). *Pauses and communication strategies in second language speech*. (ERIC Document ED 426630). Rockville, MD: Educational Resources Information Center.

- Clark, H. H., & Fox Tree J. E. (2002). Using *uh* and *um* in spontaneous speaking. *Cognition*, 84, 73-111.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cucchiari, C., Strik, H., & Boves, L. (2000). Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *Journal of the Acoustical Society of America*, 107, 989-999.
- Cucchiari, C., Strik, H., & Boves, L. (2002). Quantitative assessment of second language learners' fluency: Comparisons between read and spontaneous speech. *Journal of the Acoustical Society of America*, 111, 2862-2873.
- Davies, A. (2003). *The native speaker: Myth and reality*. (2nd ed.). Tonawanda, NY: Multilingual Matters.
- De Bot, K. (1992). A bilingual production model: Levelt's speaking model adapted. *Applied Linguistics*, 13, 1-24.
- De Jong, N. H. & Bosker, H. R. (2013). Choosing a threshold for silent pauses to measure second language fluency. In *Proceedings of the 6th Workshop on Disfluency in Spontaneous Speech (DiSS)*, Stockholm.
- De Jong, N. H., Groenhout, R., Schoonen, R., & Hulstijn, J. H. (2013). Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior. *Applied Psycholinguistics*, Advance online publication. doi:10.1017/S0142716413000210
- De Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, J. H. (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics*, 34, 893-916.
- Dehaene, S., Dupoux, E., Mehler, J., Cohen, L., Paulesu, E., Perani, D., van de Moortele, P. F., Lehericy, S. & Le Bihan, D. (1997). Anatomical variability in the cortical representation of first and second language. *Neuroreport*, 8, 3809-3815.
- DeKeyser, R. (2001). Automaticity and automatization. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 125-151). New York, NY: Cambridge University Press.
- DeKeyser, R. (2007). *Practice in a second language: Perspectives from applied linguistics and cognitive psychology*. New York, NY: Cambridge University Press.
- Derwing, T. M., Munro, M. J., & Thomson, R. I. (2007). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics*, 29, 359-380.
- Derwing, T., Rossiter, M., Munro, M., & Thomson, R. (2004). Second language fluency: Judgments on different tasks. *Language Learning*, 54, 655-679.

- Deschamps, A. (1980). The syntactic distribution of pauses in English spoken as a second language by French students. In H. W. Dechert & M. Raupach (Eds.), *Temporal variables in speech* (pp. 271-285). The Hague, Netherlands: Mouton.
- Dörnyei, Z., & Kormos, J. (1998). Problem-solving mechanisms in L2 communication: A psycholinguistic perspective. *Studies in Second Language Acquisition*, 20, 349-385.
- Duez, D. (1985). Perception of silent pauses in continuous speech. *Language and Speech*, 28, 377-389.
- Ejzenberg, R. (2000). The juggling act of oral fluency: A psycho-sociolinguistic metaphor. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 287-314). The University of Michigan Press: Michigan.
- Ferreira, F. (1993). Creation of prosody during sentence production. *Psychological Review*, 100, 233-253.
- Ferreira, F. (2007). Prosody and performance in language production. *Language and Cognitive Processes*, 22, 1151-1177.
- Field, A. P. (2005). Intraclass Correlation. In B. S. Everitt & D. C. Howell (Eds.), *Encyclopedia of Statistics in Behavioral Science* (Volume 2, pp. 948-954). Chichester: Wiley.
- Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, 21, 354-375.
- Freed, B. F. (1995). Do students who study abroad become fluent? In B. F. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 123-148). Amsterdam: John Benjamins.
- Freed, B. F. (2000). Is fluency, like beauty, in the eyes (and ears) of the beholder? In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 243-265). Ann Arbor: University of Michigan Press.
- Freed, B. F., Segalowitz, N., & Dewey, D. P. (2004). Context of learning and second language fluency in French: Comparing regular classroom, study abroad, and intensive domestic immersion programs. *Studies in Second Language Acquisition*, 26, 275-301.
- Fulcher, G. (1996). Does thick description lead to smart tests? A data-based approach to rating scale construction. *Language Testing*, 13, 208-238.
- Gass, S. M., & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Gass, S. M., & Mackey, A. (2007). *Data elicitation for second and foreign language research*. NY: Routledge.

- Ginther, A., Dimova, S., & Yang, R. (2010). Conceptual and empirical relationships between temporal measures of fluency and oral English proficiency with implications for automated scoring. *Language Testing*, 27, 379-399.
- Goffman, E. (1981). Radio talk. In E. Goffman (Ed.), *Forms of talk* (pp. 197-327). Philadelphia, PA: University of Pennsylvania Press.
- Goldman-Eisler, F. (1968). *Psycholinguistics: Experiments in spontaneous speech*. New York: Academic Press.
- Griffiths, R. (1991). Pausological research in an L2 context: A rationale, and review of selected studies. *Applied Linguistics*, 12, 345-364.
- Gut, U. (2009). *Non-native speech: A corpus-based analysis of phonological and phonetic properties of L2 English and German*. Frankfurt: Peter Lang.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2009). *Multivariate data analysis* (7th ed.). Upper Saddle River, NJ: Prentice Hall.
- Hawkins, R. R. (1971). The syntactic location of hesitation pauses. *Language and Speech*, 14, 277-288.
- Hollich, G., & Houston, D. (2007). Language Development: From speech perception to first words. In A. Slater & M. Lewis (Eds.), *Introduction to Infant Development* (pp. 170-188). New York, NY: Oxford University Press.
- Holmes, V. M. (1988). Hesitations and sentence planning. *Language and Cognitive Processes*, 3, 323-361.
- Housen, A, Kuiken, F., & Vedder, I. (2012). Complexity, accuracy and fluency: Definitions, measurement and research. In A. Housen, F. Kuiken & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency. Investigating complexity, accuracy and fluency in SLA* (pp. 1-20). Amsterdam: John Benjamins Publishing Company.
- Indiana University. (n.d.). SPEAK test rating scale. Retrieved from http://liberalarts.iupui.edu/english/index.php/academics/eap/eap_contact#rubric
- Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics*, 29, 24-49.
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice Hall.
- Kahng, J. (2012). How long should a pause be? Effects of cut-off points of pause length on analyzing L2 utterance fluency. Poster presented at Fluent Speech Workshop, Utrecht, The Netherlands.

- Kang, O., Rubin, D., & Pickering, L. (2010). Suprasegmental measures of accentedness and judgments of language learner proficiency in oral English. *Modern Language Journal, 94*, 554-566.
- Kircher, T. T. J., Brammer, M. J., Levelt, W., Bartels, M., & McGuire, P. K. (2004). Pausing for thought: Engagement of left temporal cortex during pauses in speech, *NeuroImage, 21*, 84-90.
- Koponen, M., & Riggenbach, H. (2000). Overview: Varying perspectives on fluency. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 5–24). Ann Arbor: University of Michigan Press.
- Kormos, J. (2000a). The role of attention in monitoring second language speech production. *Language Learning, 50*, 343-384.
- Kormos, J. (2000b). The timing of self-repairs in second language speech production. *Studies in Second Language Acquisition, 22*, 145-169.
- Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, NJ: Erlbaum.
- Kormos, J., & Dénes, M. (2004). Exploring measures and perceptions of fluency in the speech of second language learners. *System, 32*, 145-164.
- Kuiper, K. (1996). *Smooth talkers: The linguistic performance of auctioneers and sportscasters*. Englewood Cliffs, NJ: Erlbaum.
- Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. New York, NY: Routledge.
- Lass, N. J., & Leeper, H. A. (1977). Listening rate preference: Comparison of two time alternation techniques. *Perceptual and Motor Skills, 44*, 1163-1168.
- Lennon, P. (1984). Retelling a story in English. In H. W. Dechert, D. Möhle, & M. Raupach (Eds.), *Second language productions* (pp. 50-68). Tübingen: Gunter Narr Verlag.
- Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning, 40*, 387–417.
- Levelt, W. J. (1983). Monitoring and self-repair in speech. *Cognition, 14*, 41–104.
- Levelt, W. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W. (1999). Producing spoken language: A blueprint of the speaker. In C. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 83-122). Oxford, UK: Oxford University Press.

- MacGregor, L. J. (2008). *Disfluencies affect language comprehension: Evidence from event-related potentials and recognition memory* (Doctoral dissertation). Retrieved from Edinburgh Research Archive. (<http://hdl.handle.net/1842/3311>)
- Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word, 15*, 19–44.
- Martin, J. G., & Strange, W. (1968). The perception of hesitation in spontaneous speech. *Perception and Psychophysics, 3*, 427-438.
- Mora, J. C., & Valls-Ferrer, M. (2012). Oral fluency, accuracy, and complexity in formal instruction and study abroad learning contexts. *TESOL Quarterly, 46*, 610-641.
- O'Brien, I., Segalowitz, N., Freed, B., & Collentine, J. (2007). Phonological memory predicts second language oral fluency gains in adults. *Studies in Second Language Acquisition, 29*, 557–582.
- Opitz, B., & Friederici, A. D. (2003). Interactions of the hippocampal system and the prefrontal cortex in learning language-like rules. *NeuroImage, 19*, 1730-1737.
- Pawley, A., & Syder, F. (2000). The one clause at a time hypothesis. In H. Riggenbach (Ed.), *Perspectives on fluency* (pp. 163–191). Ann Arbor: University of Michigan Press.
- Perani, D., Paulesu, E., Galles, N. S., Dupoux, E., Dehaene, S., Bettinardi, V., Cappa, S. F., Fazio, F. & Mehler, J. (1998). The bilingual brain: proficiency and age of acquisition of the second language. *Brain, 121*, 1841-1852.
- Raupach, M. (1987). Procedural learning in advanced learners of a foreign language. In J. A. Coleman & R. Towell (Eds.), *The advanced language learner* (pp. 123-155). London: CILT.
- Reich, S. S. (1980). Significance of pauses for speech perception. *Journal of Psycholinguistic Research, 9*, 379-389.
- Riazantseva, A. (2001). Second language proficiency and pausing. *Studies in Second Language Acquisition, 23*, 297-526.
- Riggenbach, H. (1991). Towards an understanding of fluency: A microanalysis of nonnative speaker conversation. *Discourse Processes, 14*, 423-441.
- Roberts, B., & Kirsner, K. (2000). Temporal cycles in speech production. *Language and Cognitive Processes, 15*, 129-157.
- Rossiter, M. J. (2009). Perceptions of L2 fluency by native and non-native speakers of English. *Canadian Modern Language Review, 65*, 395-412.

- Schmidt, R. (2000). Forward. In H. Rigggenbach (Ed.), *Perspectives on fluency* (pp.v-viii). Ann Arbor: University of Michigan Press.
- Schnadt, M. J. (2009). Lexical influences on disfluency production (Doctoral dissertation). Retrieved from Retrieved from Edinburgh Research Archive. (<http://hdl.handle.net/1842/4424>)
- Segalowitz, N. (2000). Automaticity and attentional skill in fluent performance. In H. Rigggenbach (Ed.), *Perspectives on Fluency* (pp. 200-219). Ann Arbor, MI: University of Michigan Press.
- Segalowitz, N. (2003). Automaticity and second languages. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 382-408). Oxford, UK: Blackwell.
- Segalowitz, N. (2010). *Cognitive bases of second language fluency*. New York: Routledge.
- Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition*, 26, 173–200.
- Segalowitz, N., & Hulstijn, J. (2005). Automaticity in bilingualism and second language learning. In F. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistics approaches* (pp 371-388). Oxford, UK: Oxford University Press.
- Seidl, A., & Cristià, A. (2008). Developmental changes in the weighting of prosodic cues. *Developmental Science*, 11, 596-606.
- Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.
- Skehan, P. (2003). Task based instruction. *Language Teaching*, 36, 1–14.
- Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, 30, 510-532.
- SPSS Inc. (2008). SPSS Statistics for Windows, Version 17.0. Chicago: SPSS Inc.
- Stemler, S. E., & Tsai, J. (2007). Best practices in interrater reliability: Three common approaches. In J. W. Osborne (Ed.), *Best practices in quantitative methods* (pp. 29–49). Thousand Oaks, CA: Sage Publications.
- Sugito, M. (1990). On the role of pauses in production and perception of discourse. *Proceedings of the 1st International Conference on Spoken Language Processing, Japan*, 1, 513-516.
- SyllableCount.com. (n. d.). Syllable counter [online software]. Available from <http://www.syllablecount.com>.
- Taboada, M. (2006). Spontaneous and non-spontaneous turn-taking. *Pragmatics*, 16, 329-360.

- Tavakoli, P. (2011). Pausing patterns: Differences between L2 learners and native speakers. *ELT Journal*, 65, 71-79.
- Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239–276). Amsterdam: John Benjamins.
- Towell, R. & Dewaele, J.-M. (2005). The role of psycholinguistic factors in the development of fluency amongst advanced learners of French. In J.-M. Dewaele (Ed.), *Focus on French as a foreign language*. Tonawanda, NY: Multilingual Matters.
- Towell, R., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, 17, 84-119.
- Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effects of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition*, 28, 1-30.
- Ullman, M. T. (2001). The neural bases of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition*, 4, 105-112.
- Ullman, M. T. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition*, 92, 231-270.
- Ullman, M. T. (2005). A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model. In C. Sanz (Ed.), *Mind and context in adult second language acquisition: Methods, theory, and practice* (pp. 141-178). Washington, DC: Georgetown University Press.
- Ullman, M.T. (2013). The declarative/procedural model of language. In H. Pashler (Ed.), *Encyclopedia of the Mind* (pp. 224-226). Los Angeles: Sage Publications.
- Wood, D. (2010). *Formulaic language and second language speech fluency: Background, evidence and classroom applications*. London: Continuum.