Linguistic Research 31(3), 465-490
DOI: 10.17250/khisli.31.3.201412.003

Quantifying English fluency in Korean speakers' read-aloud and picture-cued storytelling speech*

Yongeun Lee
(Chung-Ang University)

Yongeun Lee. 2014. Quantifying English fluency in Korean speakers' read-aloud and picture-cued storytelling speech. *Linguistic Research* 31(3), 465-490. Previous studies of quantifying oral proficiency of Korean learners of English have demonstrated that temporal aspects of the learners' speech (e.g., 'rate of speech') are strongly correlated with their perceived fluency evaluated by native English-speaking raters (Choi, 2004; 2005). However, these results have been primarily based on speech elicited from contexts where Korean learners of English were asked to read simple passages or orally complete partial dialogues presented in written forms. The current study investigated whether and to what extent the previously reported findings can be generalized to different speech contexts, i.e., particularly to a more spontaneous speech. To this end, this study elicited picture-cued storytelling utterances produced by Korean learners of English and compared the results from the more spontaneous speech with the comparable results from read speech by the same speakers. The current findings indicate that perceived fluency of Korean speakers' speech can vary as a function of the modalities in which their speech produced. In addition to the types of task that the Korean speakers carried out, the present findings also show that the protocols/instructions by which the speech materials are elicited can also influence perceived fluency. Implications of the findings for objective assessments of oral fluency of Korean learners of English are discussed. **(Chung-Ang University)**

Keywords Temporal variables, Perceive fluency, Read speech, Spontaneous speech, Picture-cued storytelling, L2 speech

1. Introduction

One of the well-known differences between L1 and L2 is that producing utterances in L2 involves a considerable amount of conscious efforts (e.g., in the

* I am grateful to Sejin Oh, Seokhyung Lee and Jeongyoon Kim for their assistance with data collection. Comments from two anonymous reviewers are also greatly appreciated. All remaining errors are my own.

form of greater demand on self-monitoring). One of the important consequences of this is considerable slowing down of the speed at which L2 is delivered compared to L1 (Wiese, 1984). Since the relative rapidity with which L2 utterances are produced is quite often thought to reflect the general language command a speaker has over the foreign/second language s/he acquires, oral fluency in this sense is considered to be an important component of the general proficiency of L2 speakers. In this regard, it is not surprising that a significant body of past studies on oral fluency in the realm of L2 acquisition has focused on examining the temporal aspects (e.g., 'rate of speech', 'frequency of silent/filled pauses') exhibited by the learners' speech production.

Concerning the major goal of the current study, one of the primary interests in investigating these temporal characteristics of L2 speech comes from the observation that these objective (and thus quantifiable) aspects of learners' speech are quite strongly correlated with their perceived fluency evaluated by raters of the target language (Cucchiarini, Strik, & Boves, 2002; Cucchiarini, van Doremalen, & Strik, 2010; Lennon, 1990; Rekart & Dunkel, 1992; Riggenbach, 1991, Trofimovich & Baker, 2006). This finding is important for the field of L2 speech acquisition in that it not only sheds important light on linguistic factors affecting the perceived fluency of L2 speech but also provides useful inputs for areas that attempt to develop more cost-effective and less time-consuming tests of L2 oral fluency. In this context, in a series of work that investigated the feasibility of applying the automatic speech recognition technology in quantifying oral proficiency of Korean learners of English, Choi (2004, 2005) also reported that the fluency scores assigned by computers (that primarily make use of the temporal aspects of Korean learners' English speech) showed a strong correlation with comparable scores assigned by human (native-English speaking) raters.

To the extent that there is the possibility that objective fluency scores based mostly on the temporal aspects of learners' speech may obviate the need for human raters' (often labor-intensive and costly) assessments of Korean speakers' English fluency, it is important to further study the extent to which these findings can be generalized across different speech contexts. In this regard, it is worthy to note that most of the previously reported results were primarily based on speech gathered from contexts where Korean learners of English were asked to read simple passages or orally complete partial dialogues presented in written forms (Choi, 2004, 2005).

The primary goal of the present study is, thus, to explore the question of whether the magnitude of correlation between temporal measures and perceived fluency can vary as a function of the types of task the speakers are engaged in. More specifically, following Cucchiarini et al. (2002) and Cucchiarini et al. (2010), in this study we investigated whether and to what extent the objective measures that have been known to affect perceived fluency in read speech could be extended to more spontaneous speech, i.e., a picture-cued storytelling utterances produced by Korean learners of English.

2. Method

2.1 Speakers

Recordings from a total of 46 Korean learners of English (24 females, 22 males) were used in this study. They were all undergraduate students with various majors and at the time of recording they were attending a summer intensive English course offered by the university that they were attending. As part of requirements for the course, they took a computer-based English speaking test which was developed and administered by their university. The data analyzed in this study came from their recorded answers in this speaking test. During the test, the students saw a total of five questions, consisting of (i) a short passage reading task, (ii) a directed response task, (iii) a picture description task, (iv) a picture-cued story telling task, and finally (v) providing an opinion task. For the purpose of this study, we extracted two out of the students' five responses, i.e., a short passage-reading task and a more spontaneous speech, namely the picture-cued story telling task (more information on the contents of these two tasks are given in the text below). We distributed to the students a consent form immediately after the test to get their permission to use their recorded speech for a research purpose. In order to meet the goal of the current study, among the participants in the test, a total of 46 speakers were selected so that we could get two English proficiency groups, i.e., high-level vs. low-level English proficiency groups.

In operationalizing the English proficiency of the speakers at the two levels, we used two independent speaking test scores of the students', namely, one from the university speaking test and the other from their Oral Proficiency Interview-Computer (OPIc) test administered by the ACTFL (the students were given a chance for free of charge to take an OPIc speaking test at the conclusion of the course). The high-level group consisted of students (n=19) who obtained 16 points (out of maximum 20) from the university-developed speaking test and achieved either intermediate-High or advanced-Low level from the OPIc. The low-level participants (n=27), on the other hand, obtained 10 points (out of maximum 20) from the former test and achieved either intermediate-Low or novice-High from the latter test.

2.2 Speech material

Two types of speech material were prepared to elicit Korean speakers' production of read and spontaneous English utterances. As mentioned above, read speech was elicited by asking the participants to read aloud a short paragraph written in English. To this end, a total of three short passages were prepared and each of the speakers read one of them during the speaking test. The passages were of the similar difficulty, containing on average 8 sentences with approximately 500 words in total (see Appendix 1 for an example). For this speaking exam question the speakers were allowed 30 seconds to finish reading a passage. The recording (done as part of the speaking test) took place in a computer lab where the participants sit in front of a computer screen on which the passage appeared as a prompt. Their speech was recorded using a head-worn microphone and were saved online in the wav format to their local computer hard disk. It is important to note that they were given 45 seconds to prepare prior to reading the target passage.

More spontaneous speech was elicited by having the participants orally produce a short story using a series of hand-drawn cartoon-fashion pictures (see Figure 1 below). For this exam question, they were instructed that they should make a story in as much as detail, making use of the pictures that appeared on the computer screen as prompts. They were given 45 seconds to prepare and 60 seconds to finish. We prepared three sets of pictures and they were proportionally assigned to the speakers.

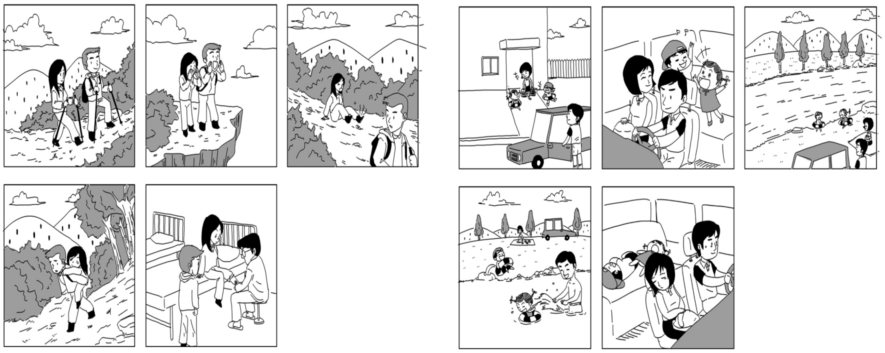


Figure 1. Examples of pictures used in the picture-cued story telling task in the speaking test developed by the students' university (Chung-Ang University holds copyright of these pictures)

2.3 Fluency raters

A group of raters of the speakers' oral fluency consisted of three native English speakers (2 male, 1 female). They were all experienced instructors teaching undergraduate level English speaking courses with more than four years of teaching experiences. The recorded speech files (a total of 92 speech files = 46 speakers x 2 types of speech) were proportionally assigned to the three raters, meaning that each of the three raters assessed fluency for about a third of the entire recordings. No single rater evaluated the same speaker's read and spontaneous speech at the same time. An attempt was made to ensure that read and spontaneous speech were evenly distributed to the three raters. Approximately a half of the entire recorded speech items (i.e., a total of 41 speech samples: 21 read and 20 spontaneous) were scored by all of the three raters so that the rating reliability across the raters can be checked later.

A review of previous studies reveals that studies differ considerably from each other in terms of the specifics of the rating rubric that they used when it comes to assessing oral fluency of L2 speakers. Some studies (Cucchiarini et al., 2010) asked raters to assess fluency using relatively finer dimensions (e.g., used separate subcategories such as speech rate, pronunciation quality, and lexical diversity, etc). However, in this study following the procedure adopted by previous studies such as Trofimovich and Baker (2006) which is most relevant to the current study, we decided that no specific instructions were given to the raters in regard how to assess fluency of their speakers, other than the requirement that their job is to assess the overall fluency of the speech and that their fluency rating score must range from 1 to 10 (10 being the most fluent). In order to assist the raters to get a rough idea about how the speech items that they were going to evaluate sound like, they were asked to rate four practice speech samples (independent from the target speech items) prior to their evaluating the target speakers.

3. Quantifying oral fluency objectively

In order to quantify fluency objectively, three steps were taken. First, all of the recorded speech items were orthographically transcribed using regular English orthography. Second, the orthographic transcriptions along with their speech files were submitted to an automatic speech recognizer to obtain rough phonemic transcriptions thereof. Finally, we then constructed a small speech database that contains the speech files, phonemic transcriptions thereof, and the alignment between the speech waves and phonemic transcriptions. A compilation of a speech database was done in order to analyze the recorded speech in terms of some representative temporal variables that are widely known to affect perceived fluency. We discuss the three steps in some detail in the following sections.

3.1 Orthographic transcription

A team of three research assistants with graduate level phonetics background initially transcribed the entire recorded speech files using regular English orthography. Each of the research assistants transcribed approximately a third of the speakers' read and spontaneous speech. For this, the transcribers relied on a transcription guideline reported in Van Engen, Baese-Berk, Baker, Choi, Kim, and Bradlow (2010). The most important principle underlying this transcription guideline is that all speech items produced, including broken and repeated words as well as various types of hesitations/filled pauses (e.g., 'uh', 'um'), are to be orthographically transcribed as exactly they were pronounced. This amounts to saying that the three transcribers did not make any corrections to any non-lexical speech disfluencies or grammatical errors found in the non-native speakers' speech productions. After the initial round of the transcriptions, all of the transcribers including the current author met several times to listen to all of the speech files together and critically evaluated the orthographic transcriptions in order to resolve any transcription discrepancies among the transcribers.

3.2 Forced alignment between phonemic labels and the speech signal

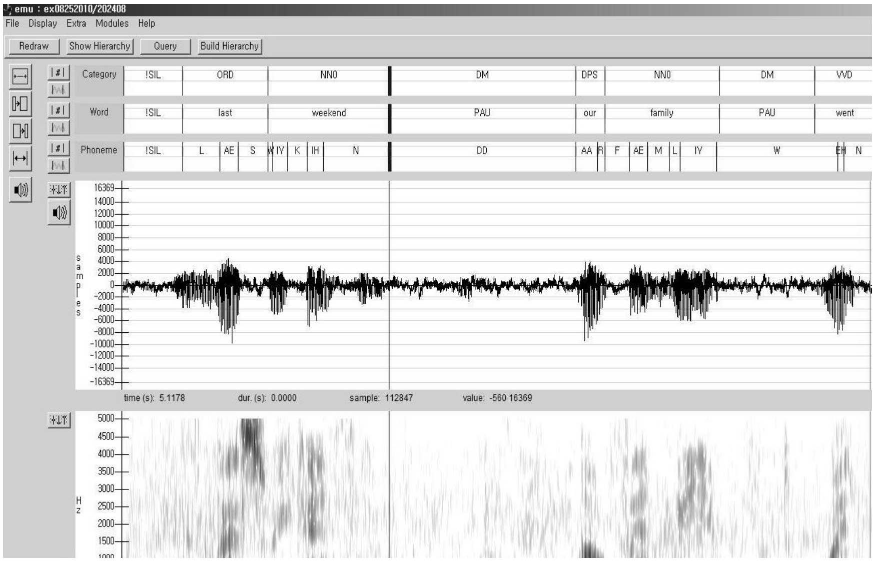
The resulting orthographic transcriptions were then saved as plain text files and these text files along with their corresponding speech files were subsequently submitted to a speech recognizer to automatically obtain rough phonemic transcriptions thereof. To this end, we used the Penn Phonetics Lab Forced Aligner (Yuan & Liberman, 2008). As is the case with other comparable speech recognition tools, this particular speech recognition toolkit takes as input the recorded speech files (in the wav format) and their corresponding orthographic transcriptions, which then subsequently returns as output Praat TextGrid files containing the rough phonemic transcriptions of the speech files. Two independent research assistants with graduate level phonetics training checked the accuracy of the phonemic transcriptions as well as the alignments between the speech signals and their phonemic transcriptions. For this task, the two research assistants were asked to pay a particular attention to the boundaries between speech and non-speech signals (i.e., silent pauses as well as filled pauses) and to check for any major phonemic transcription errors returned by the automatic speech recognizer. In regard to this, we took the special care in distinguishing the frication noises accompanied by fricatives/affricates such as /s/ or /ʧ/ from regular background noises that were present in the speech files due to the background noises. There were some considerable background noises in the speech files since the recordings took place in a speech lab with multiple talkers in it.

We found that the overall accuracy of the Penn Phonetics Lab Forced Aligner appeared to be of a reasonably good quality, at least in calculating the primary quantitative measures that are at issue in this study, namely the number of phonemes, the frequency and the length of silent/filled pauses. We note in passing, however, that since the speech recognizer was trained by a corpus based on speech produced by native English speakers, there must be some substantial numbers of errors including (but not limited to) recognizing typical phonemic errors that Korean speakers are known to make. But we feel that we are justified to use this tool for this study since our primary focus is not analyzing the exact acoustic properties of the non-native speakers' speech. In this sense, we believe that this shortcoming does not jeopardize the reliability of the results reported in this study. In fact, we believe that this study has an reliability advantage over previous studies in at least identifying pauses and calculating the number of phonemes in that the current analysis was done automatically and thus quite consistently with the help of a computer program.

3.3 Constructing a small speech database

As a final preparation step for analysis, using all of these resources gathered in the previous two steps, we constructed a small speech database which contains the speech signals, the phonemic transcriptions thereof, and the alignment between these two pieces of information (along with the word forms and their part of speech information). The construction of a speech database was done with the help of the EMU Database tool (Harrington, 2010). A snapshot of the final product is shown in Figure 2.

We decided to use this tool primarily because it permitted us to easily query the database to find the relevant information that we needed for the purpose of this study (e.g., the number of phonemes produced in a given utterance by a particular speaker). Among many other quite useful functions, the EMU tool allows one to read the phonemic transcriptions into the R statistical analysis tool (R Core Team, 2012) in the form of (what is referred to as) 'a segment list'. The tool also enables one to read Praat TextGrid files into R as a segment list through the Praat-EMU interface.



time (s): 5.1178 dur. (s): 0.0000 sample: 112847 value: -560 16369

Figure 2. A screen-shot image of the speech database constructed using the EMU tool

An example of the actual queries that we performed is given in (1) below. The meaning of this query (> emu.query ("passageReading", "202402", "Phone !=sp")) is: first, search the database named "passageReading" (first argument), second, find the utterance "202402" (second argument) contained in the "passageReading" database, third and finally, list all (phonemic) labels that are not marked as silent pauses, i.e., "sp" (third argument). As can be seen, entering the command returned 303 such cases that are not marked as "sp" in the utterance, which we considered to be the estimated number of phonemes produced in this particular utterance. One can also see that the query also returned the 'start' and 'end' time point of all phonemic labels (note that the labels are not in IPA but in SONIC, an ASCII-based transcription that is standard in automatic speech recognition research which can be read by computers). Thus, for example, the duration of the vowel "AE" (IPA /æ/) is about 40 msecs. (i.e., 401 minus 361). This way, we were able to calculate, for example, the frequency and duration of silent pauses consistently and automatically.

(1) Example of query performed using the EMU-R interface

```
> emu.query ("passageReading", "202402", "Phone !=sp")
```

```
    moving data from Tcl to R
    Read 303 records
    segment list from database: passageReading
    query was: Phone !=sp
           labels  start     end       utts
1          AE      361.67    401.58    202402
2          M       401.58    511.33    202402
3          F       511.33    541.26    202402
4          IH      541.26    571.20    202402
..........................
303         L      34833.30 34863.20 202402
```

3.4 Objective temporal measures

Previous studies investigating the relationship between temporal aspects of non-native speech and its perceived fluency have identified a variety of temporal variables that potentially contribute to perceived fluency. A recent review by Kormos (2006) provides a comprehensive list of the measures of fluency widely used in the field, ranging from temporal aspects of speech production to more phonological aspects of fluency and more recently to the use of formulaic speech. Although it would be ideal to take as many measures of fluency as possible into consideration, for the sake of feasibility of the current analysis, we decided to adopt the following measures as they have been identified as the most strong predictors of fluency by the majority of the previous studies.

- Speech rate: This measure refers to the number of phonemes produced in a given utterance divided by the time (unit: second) it takes for the speaker to produce the utterance (including pause time).
- Articulation rate: This measure is the same as the speech rate except that the duration of pause is excluded. Thus articulation rate is the time that a speaker spent speaking when the speaker produced an utterance.
- Phonation-time ratio: This is the percentage of time spent speaking as a proportion of the time it took to produce the utterance.
- Mean length of runs: This measure refers to the average number of

phonemes produced between silent pauses of 0.20 seconds and above. Note that in identifying silent pauses, juncture pause (pause between two utterances) was marked so that speech duration was calculated with reference to the onset of the first word and the offset of the last word of every utterance.

- The mean length of silent pauses: This measure refers to the total length of silent pauses over 0.20 seconds divided by the total number of silent pauses above 0.20 seconds.
- The duration of silent pauses per minute: This is the duration of silent pauses expressed in seconds divided by the time the speaker spent speaking and multiplied by 60. We calculated duration of silent pauses relative to utterance length since unlike the read speech where speakers produced roughly the same amount of speech, in the picture-cued story telling task, participants produced variable utterances lengths, that is, utterance fragments of different length had to be compared.
- The number of silent pauses per minute: This measure is the same as "duration of silent pauses per minute" except that the numerator is the number of silent pauses.

4. Results

4.1 Fluency ratings

All together, a total of 41 speech samples (21 read and 20 spontaneous) were rated by all of the three raters for interrater reliability check. The reliability coefficients (Cronbach's alpha) were 0.67 and 0.72 for read and spontaneous speech, respectively. Since an alpha of 0.7 and over is widely accepted as an indication of acceptable reliability (George & Mallery, 2003; Kline, 2000), we believe that the interrater reliability pooled over the current raters is reasonably high. Thus, in the following we report the means and standard deviations of the fluency rating scores pooled over across the three raters.

Table 1. Means and standard deviations of fluency rating scores for read-aloud vs. story-telling speech of speakers of high vs. low proficiency level

	Read-aloud task			Picture-cued storytelling task		
	Low	High	All	Low	High	All
	mean (s.d.)	mean (s.d.)	mean (s.d.)	mean (s.d.)	mean (s.d.)	mean (s.d.)
Fluency scores	6.57 (1.36)	8.25 (1.41)	7.49 (1.61)	4.73 (1.61)	6.86 (1.25)	5.80 (1.79)

Overall, we found evidence that factors such as speakers' English proficiency level and the types of oral tasks that they are engaged in influence raters' assessments of speakers fluency. First, with regard to fluency scores as a function of speakers' proficiency level, it was the case that fluency scores increase as the proficiency level changes from low to high. This held true for both of the read-aloud and the storytelling task. Secondly, when pooled over the two proficiency levels, the speakers obtained lower fluency scores when they were engaged in the more spontaneous picture-cued storytelling task than in the read-aloud task. Specifically, as shown in Table 1, the grand mean fluency score of the story-telling task (pulled over the two proficiency levels) was substantially lower (5.80) compared to that of the read-aloud task (7.49). The fact that the current speakers obtained much lower fluency scores in the more spontaneous speech was especially evident from the fact that the mean fluency score of the high-level speakers from their storytelling task was 6.86, a number that is quite close to the mean fluency score of 6.57 obtained from the low-level speakers' read-aloud task.

An obvious account of this result is that, compared to the read-aloud task, producing a short but coherent story based on a series of pictures should have been more difficult for the current participants across the board. That is, since the task of making a story involves not only orally producing English speech but also selecting proper words and subsequently organizing them into proper sentences online, spontaneously produced storytelling speech should have been cognitively more demanding on the part of the current Korean talkers. To make the task even more difficult, the storytelling task occurred under a speaking test setting, which could have burdened even more processing load on the Korean speakers' side. All of these are highly likely to have reduced fluency scores across the board. Indeed, a review of past studies reveals that tasks that involve more processing load indeed heavily influence fluency scores. Cucchiarini et al. (2002) and Grosjean (1980), for example, reported that the more speech is spontaneous the more often the speech is associated with lower speech rate, which contributes to lowering fluency scores especially for L2 learners. According to them, the lower speech rate was attributed to lower articulation rate and especially to lower phonation/time ratio, shorter runs, and longer pauses. Analyses of the quantitative measures of fluency obtained in this study will be presented in the next section to see whether this was indeed the case for the current speakers.

4.2 Quantitative measures of fluency

Table 2 below contains means and standard deviations for all of the variables for fluency considered in this study (following Table IV in Cucchiarini, et al., 2002: pp. 2868). The table is firstly divided into two primary columns as a function of the mode of speech that the speakers were engaged in, i.e., read (R) and spontaneous speech (S). Secondly, each of the two primary modes of speech in turn contains mean values obtained from the low-level (column 1 and 4) and high-level speakers (column 2 and 5) as well as the grand means (column 3 and 6) calculated across the two proficiency groups of speakers within each of the two modes of speech. Lastly, in order to give readers an indication of how the mean values change across the two modes of speech, the ratio between the mean values from read speech and the corresponding values from the spontaneous speech is given in column 7. The ratios were obtained by dividing the average values in column 6 by the average values in column 3.

In the table, the three top rows, i.e., 'number of phoneme', 'speech time without pauses', and 'speech time with pauses', present an overview of the amount of speech material that this study analyzed. From 'speech time including pauses' one can see that on average the present speakers produced about a minute length of spontaneous speech (52.8 sec.) as opposed to about half a minute length of read speech (31.8 sec.). Given that the speakers were instructed to complete their tasks within 60 sec. and 30 sec. for the story-telling and read-aloud task respectively, these two average values from 'speech time including pauses' indicate that the speakers managed to finish the tasks approximately within the alloted time (although the speakers spent about on average 7 seconds short of the allotted time for the storytelling task).

At this point, it is worth noting that despite the numerical difference in total durations that were taken to complete the two tasks, the fact that the two modes of speech do not differ so much in terms of the means of 'duration of speech excluding pauses' (31.1 sec vs. 26.7 sec) clearly indicates that the current speakers produced much longer silent pauses and/or produced pauses more frequently when they were involved in the storytelling task than in the read-aloud task. In addition, although on average it took much longer (almost twice longer) for the speakers to finish the storytelling task (52.8 sec.) than the read-aloud task (31.8 sec.), the speakers produced much fewer number of phonemes in the former (10) than in the latter (31) task. This corroborates the fact that the speakers produced longer and/or more silent pauses when they were engaged in the spontaneous speech than in the read speech. This finding is thus in support of the findings in previous studies that one of the major characteristics of the spontaneous speech by non-native speakers of English is the frequent occurrences of speech disfluency, including silent and filled pauses.

Table 2. Means and standard deviations for all of the fluency variables in this study

Type of Speech	Read-aloud task (R)			Storytelling task (S)			
	1	2	3	4	5	6	7
Measures of Fluency	Low [n=27] mean (s.d.)	High [n=19] mean (s.d.)	all-R [n=46] mean (s.d.)	Low [n=27] mean (s.d.)	High [n=19] mean (s.d.)	all-S [n=46] mean (s.d.)	S/R ratio. col.6/ col.3
Number of phonemes	287.11 (39.9)	289.89 (11.7)	288.26 (31.2)	208.81 (67.1)	248.36 (55.3)	225.15 (64.9)	
Speech time excluding pauses (sec.)	27.5 (3.8)	25.5 (1.5)	26.7 (3.2)	31.1 (8.2)	31.1 (7.6)	31.1 (7.8)	
Speech time including pauses (sec.)	33.2 (6.1)	29.8 (2.6)	31.8 (5.2)	56.4 (10.5)	47.7 (7.9)	52.8 (10.0)	
Articulation rate	10.6 (2.1)	11.36 (0.7)	10.92 (1.7)	6.77 (1.4)	8.11 (1.2)	7.32 (1.5)	0.67
Rate of speech	8.91 (2.0)	9.79 (0.8)	9.27 (1.7)	3.72 (1.0)	5.30 (1.3)	4.37 (1.4)	0.47

Phonation/ time ratio	83 (6.2)	86 (4.1)	85 (5.5)	55 (10.3)	65 (12.3)	59 (12.1)	0.69
Mean length of runs	31 (12.7)	32 (8.2)	31 (10.9)	8 (1.8)	12 (4.4)	10 (3.6)	0.32
Mean length of silent pauses (unit: sec.)	0.54 (0.02)	0.43 (0.91)	.050 (0.16)	1.08 (0.46)	0.78 (0.28)	0.96 (0.41)	1.92
Duration of silent pauses per min.	10 (3.7)	8 (2.5)	9 (3.3)	27 (6.2)	21 (7.4)	24 (7.3)	2.66
Number of silent pauses per min.	0.019 (0.004)	0.019 (0.004)	0.018 (0.004)	0.027 (0.006)	0.026 (0.004)	0.026 (0.005)	1.44

In regard to the rest of the fluency variables, the S/R ratio values in column 7 also indicate that the mode of speech factor influenced the fluency variables analyzed in the current study. More specifically, 'articulation rate', 'rate of speech', 'phonation/time ratio', and 'mean length of runs' decreased quite noticeably as the speech mode changes from read to spontaneous. This is particularly true for 'rate of speech', which is approximately halved (0.47) and also for the number of phonemes produced between two silent pauses (i.e., length of runs), which is more than halved (0.32). The opposite pattern emerged from 'mean length of silent pauses' and 'duration of silent pauses per minute'. Notably, the duration of silent pauses is more than doubled (2.66) when the mode of speech changes from read to spontaneous speech.

The results taken together again suggest that the current speakers talked in a much slower manner, produced relatively shorter phrases/sentences and longer silent pauses in spontaneous than in read speech. Overall, the quantitative measures, thus, indicate that the current speakers appeared to be less fluent when they were engaged in the storytelling task, if fluency can be broadly defined in terms of temporal aspects of speech, more specifically in terms of the relative degree of rapidity involved in delivering speech (Lennon, 1990). As mentioned above, previous studies have suggested that when non-native speakers were engaged in more cognitively demanding task their speech tend to be less fluent than when they were involved in less cognitively demanding task (Cucchiarini, et al., 2002; Grosjean, 1980). This is likely to be what is happening in the current study as well, since the task used in the current spontaneous speech was to orally present a short but coherent story in English, making use of a series of cartoon pictures online (as opposed to simply reading aloud a given short paragraph).

Next, the change of the quantitative measures of the fluency variables as a function of the participants' English proficiency level can be examined by comparing columns 2 and 3 within 'R' and columns 5 and 6 within 'S', respectively. First, it is interesting to see that within 'R', there was only a minor increase or almost no change in 'articulation rate', 'rate of speech', 'phonation/time ratio', and 'mean length of runs' as the proficiency level changes from low to high. Practically, the same pattern is observed when the length and the frequency of silent pauses are considered. This means that the low-level Korean speakers had little problem in reading aloud the passages that were given to them. Overall, the two proficiency groups thus seem not to have diverged from each other when they were in the read-aloud task. This in turn implies that the fluency variables adopted in this study may not serve as viable factors in predicting the proficiency level of Korean speakers at least in the context of read-aloud tasks.

In the context of storytelling, on the other hand, the difference between the less and the more proficient groups was more prominent. That is, compared to the read speech there was a greater increase in 'articulation rate', 'rate of speech', 'phonation/time ratio', and 'mean length of runs' as the proficiency level changes from low to high. Likewise, there was a greater decrease in terms of 'length of silent pauses' in the spontaneous speech as the proficiency level changes (although, interestingly, this pattern is not observed in terms of the frequency of silent pauses per minute). Note that in this study both 'articulation rate' and 'rate of speech' are measures relative to the number of phonemes. In this regard, it is interesting to note that in spontaneous speech, even though the high-proficiency group on average produced much greater number of phonemes (248) than the low-proficiency group did (208), the former group's mean 'articulation rate' and 'rate of speech' values are still greater than the latter group's. This suggests that in spontaneous speech the high-proficiency group indeed presented their stories orally a lot faster with shorter pauses than the low-proficiency group. In sum, the current results strongly suggest the possibility that the fluency variables adopted in this study may serve as more efficient measures in predicting the degree of fluency of the present Korean speakers when they were engaged not in reading pre-arranged scripts but in more spontaneous speech. This possibility will be explored in detail in the next section.

4.3 Relation between temporal measures and perceived fluency ratings

This section reports our findings that are most relevant to the primary goal of this study, namely the extent to which the objective temporal measures of fluency are correlated with the subjective perceived fluency scores assigned by native English-speaking raters. Following previous studies (Cucchiarini et al., 2002), the correlations between the mean rating scores and the objective fluency measures were quantified using Pearson's *r* correlation coefficients. Table 3 below lists all correlation coefficients calculated in this study and they are organized first by the two type of tasks shown in the two columns, referred to as 'all-R' (read-aloud) vs. 'all-S' (storytelling). Within each task, in turn, correlations are reported for the two proficiency levels separately.

When considering all correlations, we found, as expected, that 'articulation rate', 'rate of speech', 'phonation/time ratio', and 'mean length of runs' are in general positively correlated, whereas 'mean length of silent pauses', 'duration of silent pauses per min', and 'number of silent pauses per min.' are in general negatively correlated with the fluency scores assigned by the native speaker raters (except the finding from the low-level speakers in the storytelling setting where the number of silent pauses per min. was actually positively correlated with the fluency rating). This overall finding is consistent with the reports from most of the previous studies. That is, in general the faster the speaker produces his/her utterances, the higher their fluency scores tend to be. It appears to be also the case that the less one's speech contains disfluency, the higher their fluency scores are.

4.3.1 Read-aloud speech

In spite of this overall result that is in general agreement with previous findings, we also found some patterns that diverge from the previous studies. Most notably, we observed that none of the correlations from the read-aloud task were statistically significant, although in terms of the magnitude of the correlations, it appears that the correlation coefficients of 'articulation rate' and 'rate of speech' are numerically bigger than those of the other fluency measures. This result suggests that at least for the current read speech, except for the potentially minor effect of 'rate of speech', the other fluency measures which we considered did not play a significant role in determining the current raters' perception of fluency.

Table 3. Correlations (Pearson's r) between the perceived fluency scores and the temporal variables for the two proficiency levels sorted by the two modes of speech (* indicates that the correlation was significant at the 0.05 level)

Type of Speech	Read-aloud (R)			Story-telling (S)		
Measures of Fluency	Low [n=27]	High [n=19]	all-R [n=46]	Low [n=27]	High [n=19]	all-S [n=46]
Articulation rate	0.43	0.60	0.45	0.30	0.40	0.51
Rate of speech	0.49	0.55	0.51	0.72*	0.71*	0.78*
Phonation/time ratio	0.38	0.17	0.35	0.67*	0.56	0.69*
Mean length of runs	0.34	-0.04	0.17	0.25	0.59	0.56
Mean length of silent pauses	-0.25	-0.40	-0.40	-0.75*	-0.54	-0.75*
Duration of silent pauses per minute	-0.38	-0.17	-0.35	-0.67*	-0.56	-0.69*
Number of silent pauses per minute	-0.22	0.08	-0.03	0.53	-0.18	0.28

One reasonable explanation of this finding is that unlike more spontaneous speech, read speech in general (including read speech of L2 learners in particular) contain relatively little amount of speech disfluency. In fact, it is generally the case that speech disfluencies including silent and filled pauses are quite infrequent when people simply read passages. This is especially true when they are allowed to examine the passage that they are going to read beforehand. We have shown above that this was indeed the case in the current study where the speakers paused less frequently and the duration of the pauses was a lot shorter in their read vs. in spontaneous speech (see Table 2). In this regard, we believe that the way in which the read speech was collected in this study should have contributed to this. That is, as described above, the passages used in the current read-aloud task contained about 8 sentences. Importantly, the speakers were given 45 seconds to prepare prior to reading the target passage, during which we believe they should have sufficiently practiced producing the material, albeit not overtly (i.e., silently). Simply speaking, for both of the high and low level groups alike, the task of reading aloud a passage with sufficient preparation time should have allowed them to read the passages in a comfortable tempo, yielding less pauses and/or hesitations and thus significant increases in the length of fluent runs.

Given this, it is possible that in assessing fluency of the read speech the raters could have resorted to some fluency-related variables other than the typical temporal variables, i.e., rate of speech or number of speech disfluencies. Possible variables include pronunciation quality of individual segments and suprasegmental features such as the number of stressed words per unit time and appropriate pitch contour patterns (Vanderplank, 1993). The fact that the rating rubric provided to the current raters did not distinguish these fine dimensions of fluency may also have contributed to the current finding. The result from the read-aloud task thus suggests that in evaluating the predictive power of fluency measures for perceived fluency of English read speech by Korean learners, it is very important to take into consideration the precise protocol by which the read-aloud speech was collected, including the amount of preparation time allowed and the nature of the rating rubric used.

4.3.2 Spontaneous speech

The most noteworthy pattern when considering the correlations for spontaneous speech was that the magnitudes of all correlation coefficients appeared to be systematically higher for the spontaneous speech (see column 'all-S' in Table 3) than for the read speech (see column 'all-R' in the same table). This suggests that the power of the objective measures used in the current study appears to be greater in predicting perceived fluency of (at least) the current Korean speakers' spontaneous compared to read speech. Given that the fluency measures in the current study were mainly consisted of the temporal aspects of speech, this finding suggests that the degree of perceived fluency can be more reliably estimated by temporal variables when the learners' speech was produced in a more spontaneous context. This is in line with previous reports such as Cucchiarini et al., (2002) where they also reported that correlations between fluency scores and most of the temporal variables that they examined were highly significant when L2 speech was produced in a more spontaneous setting.

With regard to the temporal variables that we examined, 'rate of speech', 'phonation/time ratio', and 'duration of pauses' showed significant correlations with fluency scores when the data from the two proficiency levels were pooled together. Among them, it was 'rate of speech' that returned the largest correlation coefficients.

On the basis of this, we might infer that as long as temporal variables are concerned, the relative speed with which the speech is produced and the duration of silent pauses could serve as reasonably good predictors of perceived fluency of learners' spontaneous L2 speech. Related to this, the current result also suggests that unlike the read speech where the speakers produced little amount of silent pauses, in the current picture-cued storytelling task, our participants produced variable lengths of pauses with a greater frequency. In this regard, it is worth noting that in spontaneous speech 'mean length of runs' (i.e., the amount of phonemes spoken between two silent pauses) appeared to be less related with perceived fluency than other temporal variables, especially the above mentioned 'length of pauses'. This set of results raises the possibility that as the pause length in an utterance increases quite noticeably (actually, from 1.5 times to almost 2 times in the current study), which is typically the case when the speech changes from read to spontaneous speech, the perceived degree of fluency on the part of native English speakers decreases quite noticeably regardless of the number of phonemes that L2 learners produce. This in turn implies that in order for the Korean learners of English to obtain better scores on perceived fluency in spontaneous speech setting, producing shorter pauses might be more important than producing more phonemes in a given run, i.e., longer stretches of phrases or sentences.

In order to assess the best linear combination of the fluency variables that can be used to better predict the mean rating scores in spontaneous speech, we further analyzed the correlation data using a multiple regression analysis method. The result of this analysis showed that, as expected, 'rate of speech' accounted for the greatest amount of variance (R = 0.61, F = 66.43, df =1). 'Duration of silent pauses per minute' came in the second place (R = 0.48, F = 39.03, df = 1). We added the latter variable to the former in the stepwise procedure to see whether the combination of these two lead to significant improvements in predictive power. The increase in the amount of variance that was accounted for by this procedure appeared to be only marginal (i.e., R rose from 0.61 to 0.63, F = 35.9, df = 2).

4.3.3 Role of proficiency

With respect to the role of proficiency levels in the pattern of perceived fluency in spontaneous speech, we observed the following pattern. Firstly, within the low-level group, 'rate of speech', 'phonation/time ratio', and the two variables regarding the pauses showed significant correlations with fluency ratings. Among them, it was 'rate of speech' and 'mean length of silent pauses' that had the largest correlation coefficients. A multiple regression analysis using the fluency measures as predictors and the rating scores as the criterion was performed. The result showed that 'mean length of silent pauses' accounted for the greatest amount of variance ($R = 0.58$, $F = 28.9$, df =1), with 'rate of speech' coming in the second place ($R = 0.52$, $F = 24.44$, $df = 1$). We added the latter variable to the former in the stepwise procedure to see whether the combination of these two lead to significant improvements in predictive power. The increase in the amount of variance that was accounted for by this procedure appeared to be only marginal (i.e., R changed from 0.58 to 0.60, $F = 16.46$ $df = 2$). Secondly, within the high-level group, it was only 'rate of speech' that was statistically significant. This variable accounted for about the half of the variance ($R = 0.50$, $F = 17.46$, df =1). When the 'mean length of silent pauses' was added to the statistical model, there was little increase in the explained variance (R changed from 0.50 to 0.51, $F = 8.25$ $df = 2$).

In sum, it appears that 'rate of speech', on the one hand, seems to be the best predictor of perceived fluency for both of the high-level and low-level groups when the task is to orally produce a short story spontaneously. On the other hand, the presence or absence of a significant correlation between 'length of pauses' and fluency ratings is one major difference between the two proficiency groups. Namely, 'length of pauses' appeared to have influenced English native speakers' perception of spontaneous speech produced by the low-level group but not for the same type of speech produced by the high-level group. As mentioned above, this again implies that for Korean speakers to obtain higher fluency ratings in spontaneous speech, it seems more important for them to produce their utterances in a reasonably high tempo. Given that the spontaneous speech by the high-level group was indeed on average more rapid compared to the speed adopted by the low-level group (see Table 2 above), 'length of silent pauses' seems not to have adversely affected the former group's perceived fluency.

5. Discussion and Conclusions

The major goal of this study was to investigate the extent to which subjective fluency scores assigned by native speakers of English are correlated with major objective temporal variables calculated for read and spontaneous speech by Korean learners of English at two gross proficiency levels.

First of all, we found that objective measures of fluency can vary as a function of the speech modalities that the speakers are engaged in. If fluency can be broadly defined in terms of temporal aspects of speech (more specifically in terms of the relative degree of rapidity involved in delivering speech), the speakers appeared to be less fluent when they were engaged in the more spontaneous story-telling task than in the read-aloud task.

More importantly to the current research question, we found that the reliability of objective measures as an indicator of perceived fluency varied as a function of the type of speech that the speakers were engaged in. Specifically, although most of the previous studies have suggested that perceived fluency scores are quite reliably predicted by a limited set of objective temporal measures calculated for read speech, we obtained results that are apparently inconsistent with this. That is, we found no significant correlations between fluency scores and temporal variables in the read speech. In spite of this finding, we do not believe that this indicates that the traditionally recognized temporal variables are only weakly related to perceived fluency of read speech. Rather, we believe that the current finding merely suggests that not only the nature of the task but also the specific protocol and/or instructions whereby the read-aloud speech is gathered can make a difference. In this respect, we suppose that the current finding is just an artifact that reflects the particular way in which the current read speech was elicited. In regards to this, in the above we have discussed several potential factors that might have contributed to the discrepancy between the current and previous findings. Included to this is the fact that, in the current study, the participants were allowed reasonably sufficient time prior to reading aloud the written passage. Unlike this, in previous studies it was often the case that subjects had to read the sentences immediately after the prompts were given, meaning that they were not able to prepare themselves beforehand. This could have been particularly true if the passages had contained words that were relatively less familiar to the speakers. This might have caused them to display more speech disfluencies such as silent or filled pauses. Given that temporal variables associated with the current read speech did not apparently affect listeners' assessments of fluency, there must be some hidden factors that have influenced the current judges' assessments. Possible factors discussed above include the pronunciation quality of individual segments as well as suprasegmental features such as the number of stressed words per unit time and/or proper pitch contour patterns (Vanderplank, 1993). In our future study, we plan to investigate the role of these variables in perceived fluency of read speech by Korean learners of English.

Unlike read speech, we found that the power of the objective measures used in the current study appears to be reasonably good in predicting perceived fluency of the current Korean speakers' spontaneous speech. Under the more narrow interpretation, L2 fluency is usually defined by many researchers to be able "to talk at length with few pauses and to be able to fill the time with talk" (Kormos, 2006: pp. 155). Given this, it is not surprising that the temporal variables in this study could serve as reasonably good predictors of perceived fluency of the current speakers' spontaneous L2 speech since that was the precisely the mode in which we were able to best observe variable rate of speech and length of pauses among the speakers. This is in contrast to read speech where the speakers produced little amount of speech disfluency. Another interesting finding from the current spontaneous speech was that 'mean length of runs' (i.e., the amount of phonemes spoken between two silent pauses) appeared to be less related with perceived fluency than other temporal variables related to pauses such as 'length of pauses'. Interestingly, this was particularly true for the low-proficiency group, which suggests that pause-related variables for the low-level Korean speakers in spontaneous speech may be more important than the variables related to the number of phonemes/syllables produced in a given unit time. One hypothesis that may account for this is that when the impressionistic fluency of L2 speech is quite low on the native English speakers' part in the first place, in their subsequent subjective evaluation they may put more emphasis on how rapidly the speech that they hear is executed than on how contentful the speech is (indicated in part by the number of phonemes produced in-between two pauses). Under this interpretation, when native English-speaking judges evaluate fluency of low-level speakers' spontaneous speech it is not so much about the amount of semantic content (estimated by the length of phrases or sentences) as about a mostly performance phenomenon (Lennon, 1990). As mentioned above, an upshot of this is that when a Korean speaker attempts to deliver a brief but coherent story in English, it might be better for him/her to make the story contain shorter pauses with less phonemes than longer stretches of phrases or sentences with long pauses. In fact, this is further supported by the finding that 'rate of speech' seems to be the best predictor of perceived fluency for both high- and low-level groups when the task is to orally produce a short but coherent story spontaneously.

To conclude, our results indicate that the traditionally recognized temporal variables of speech could well relate to perceived fluency of Korean speakers. However, when estimating the magnitude of the effect of the objective measures on perceived fluency, researchers need to take special care in controlling for the way the speech is elicited, including the types of task that the speakers carry out, protocols by which the speech material is elicited, the target language proficiency the learners have, and rating rubric provided to the fluency raters.

References

Choi, Inn.-Chull. 2004. Applicability of automatic speech recognition technology to computer-based simulated oral proficiency interview. *Multimedia Assisted Language Learning* 7(2): 335-351.

Choi, Inn.-Chull. 2005. Measurability of oral fluency through ASR-based COPI. *Multimedia Assisted Language Learning* 8(2): 240-261.

Cucchiarini, Catia, Helmer Strik and Lou Boves. 2002. Quantitative assessment of second language learners' fluency: Comparison between read and spontaneous speech. *Journal of the Acoustical Society of America* 111(6): 2862-2873.

Cucchiarini, Catia, Joost van Doremalen and Helmer Strik. 2010. Fluency in non-native read and spontaneous speech. In *Proceedings of DiSS-LPSS Joint Workshop 2010*, Tokyo, Japan.

George, Darren. and Paul Mallery. 2003. *SPSS for Windows step by step: A simple guide and reference*. Boston: Allyn & Bacon.

Grosjean, Francois. 1980. Temporal variables within and between languages. In H. Dechert and M. Raupach (Eds.) *Towards a Cross-Linguistic Assessment of Speech Production.* Lang, Frankfurt, pp. 39-53.

Harrington, Jonathan. 2010. *Phonetic Analysis of Speech Corpora.* West Sussex: Wiley-Blackwell.

Kline, Paul. 2000. *The Handbook of Psychological Testing*. London: Routledge.

Lennon, Paul. 1990. Investigating fluency in EFL: A quantitative approach. *Language Learning* 3: 387-417.

R Development Core Team. 2011. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rekart, Deborah. and Dunkel, Patricia. 1992. The utility of objective (computer) measures of the fluency of speakers of English as a second language. *Applied Language Learning* 3: 65-85.

Riggenbach, Heidi. 1991. Towards an understanding of fluency: A microanalysis of non-native speaker conversation. *Discourse Processes* 14: 423-441.

Trofimovich, Pavel. and Baker, Wendy. 2006. Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition* 28: 1-30.

Van Engen, Kristin J., Baese-Berk, Melissa, Baker, Rachel E., Choi, Arim, Kim, Midam, and Bradlow, Ann R. 2010. The Wildcat corpus of native- and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and Speech* 53: 510-540.

Wiese, Richard. 1984. Language production in foreign and native languages: Same or different? In H. Dechert, D. Möhle, & M. Raupach (Eds.) *Second Language Productions*. Tübingen, Germany: Narr, pp. 11-25.

Yuan, Jiahong. and Liberman, Mark. 2008. Speaker identification on the SCOTUS corpus. *Proceedings of Acoustics 2008*, pp. 5687-5690.

Appendix 1. Example of read-aloud passages

Passengers travelling with significant luggage or young children should transfer by taxi to their continuing train at the South station. Other passengers should connect to their continuing train using the orange line subway. The station is accessible to passengers with disabilities. The best way to make sure that you receive the assistance you require at the station is to specifically request assistance when you make your reservation.

Yongeun Lee
Department of English Language & Literature
Chung-Ang University
84, Heukseok-ro, Dongjak-gu, Seoul, 156-756 Korea.
E-mail: yelee@cau.ac.kr

Received: 2014. 09. 15.
Revised: 2014. 11. 25.
Accepted: 2014. 11. 25.