Linguistic Research 36(Special Edition), 59-80
DOI: 10.17250/khisli.36..201909.003

A study on evaluating English speech rate: Focusing on professional vs. non-professional raters*

Mi-Sun Kim
(Hankuk University of Foreign Studies)

Kim, Mi-Sun. 2019. A study on evaluating English speech rate: Focusing on professional vs. non-professional raters. *Linguistic Research* 36(Special Edition), 59-80. This study is designed to find out what speeds professional evaluators give the best scores to on speaking assessments and how their speed-related assessment differs from non-expert evaluators. The test tokens are designed to avoid the experimental complications that have normally appeared in previous studies of relevance. The test materials recorded by a trained native speaker with 8 different speed ranges are free of distortion caused by computer manipulation, and they exclude other factors affecting rating results as much as possible. Three different groups of eight each participated (8 native English raters and 8 Korean raters who had official evaluation experience on an English-speaking test as professional raters, and 8 general Korean student-raters with high English proficiency as non-professional raters), with 24 raters in total, and the results show that the range of the speaking rate that draws the highest score is between 4.43 and 4.50 syl/sec on average. The assessment of extremely fast speed ranges varied slightly among the three rater groups but all three groups showed a similar tendency. **(Hankuk University of Foreign Studies)**

Keywords speech rate, speech rate preference, optimal speaking rate, speaking rate assessment

1. Introduction

The importance of speaking rate in second language (L2) speech has been one of the major issues in the L2 evaluation area. Listeners assign better scores to faster speech (Derwing et al. 2004), regard faster speech as more native-like (Kang and Rhee 2011), and consider faster speech as less accented (Trofimovich and Baker 2006; Lee 2018). However, which rate of speech gets the best score has not been properly examined in a controlled manner, especially in the context of English speaking assessment.

* I would like to thank the anonymous reviewers of this journal for their valuable comments.

The most preferred speaking rate reported in previous studies was normally around less than 5 syl/sec (syllables per second). In Foulke and Sticht (1966), the most preferred speaking rate was 207 wpm (words per minute) which is approximately 4.7 syl/sec,[1] and 200 wpm in Lass and Leeper (1977), approximately 4.50 syl/sec. Munro and Derwing (1998) reported that the most optimal rate for native speakers was 4.7 syl/sec, 4.84 in Isaacs (2008), and 4.69 in Kang (2010), which are similar to the average speech rate of native speakers.

Though it varies among studies depending on their research type, the average speaking rate for a native speaker is normally between 4 and 6 syl/sec. In Anderson-Hsieh and Venkatagiri (1994), the average speech rate of a native speaker was 5 syl/sec, 4.9 in Munro and Derwing (1998), 4.84 in Isaacs (2008), and approximately 4.91 syl/sec (originally 218 wpm) in Götz (2013).

However, there are two major issues with previous research. First, the rating results may have been biased by distorted computer manipulated materials. In some studies, researchers used artificial computer manipulation to obtain various speaking rate stimuli (Foulke and Sticht 1966; Apple et al. 1979; Munro and Derwing 1998, 2001). Additionally, due to technological limitations, a certain degree of distortion was inevitable, such that only a limited range of speech rates (normally 3 speed ranges) could be applied for the test stimuli. Second, apart from speaking rate, other important features affecting the assessment results were not excluded, as they used uncontrolled L2 utterance data (Derwing et al. 2009; Kang 2010; Kang and Rhee 2011). In other words, the reason for a low-scoring utterance, whether due to a slow rate of speech, segmental errors, or an awkward rhythm, is not clear.

In order to improve upon these limitations shown in previous studies, this study tried to exclude various factors that could affect the assessment results and minimized computer-modified manipulation to obtain natural utterances without distortion.

Also, this study selected evaluators who have experience in official evaluations to help English learners preparing for English speaking tests to find the speed of English utterance is that most preferred by professional evaluators.

---

1 syl/sec = (wpm/60)*1.35. The average word length in English consists of 1.3 ~ 1.4 syllables per word (Lehtonen 1979). Also, Munro and Derwing (1998) assumed that 1 word consists of approximately 1.35 syllables, in their calculation of word per minutes to syllable per seconds.

The same test materials applied in Kim and Jang (2016) were used; however, in this paper, the group of evaluators was limited to a group of experts, and the general group of evaluators was also limited to those with high English proficiency in order to match their English proficiency to those of professional evaluators as much as possible.

In sum, unlike the previous studies of relevance, in this study,

[1] stimuli of various speaking rates are applied;
[2] the stimuli are absent of other factors affecting the assessment results other than speech rate;
[3] professional evaluators participated.

To determine what speed of English utterance would be most preferred by professional evaluators, a total of 16 expert evaluators (8 English native speakers and 8 Korean English teachers) participated in the experiment. For comparison with a group of experts, 8 non-expert Korean student evaluators also participated. In order for evaluators to solely focus on speech rate in their assessment, speech from a trained native speaker was recorded and manipulated to be used as test material. The key purpose for manipulating the test tokens was to diversify the rates of speech that could cover most of the speed ranges that humans can produce. Since speaking rate is highly related to utterance length, the length of the test sentences was differentiated as well. Therefore, there are three major factors which are taken into consideration in this experiment: speech rate, utterance length and rater group.

2. Experiment

2.1 Participants

A total of 24 raters, 8 native English expert raters (ER), 8 Korean expert raters (KR), and 8 Korean general (non-expert) raters (KG) participated in this experiment and reported that all had normal hearing. Determining expertise in the field of assessment was not simple, so the expert groups, ER and KR

comprised only those with official evaluation experience of an English-speaking test. Additionally, in order to check the evaluation trends of the general public (ordinary people with high English proficiency who are capable of evaluating but who do not have any special qualifications or experience rating official English tests), student raters (KG) participated. Thus, ER and KR are considered to be expert evaluators, and KG is a group of non-expert evaluators.

All of the ER raters had been teaching English for 7 to 21 years, and most of the KR raters were also teaching English at a university. All KR raters hold master's or doctoral degrees in English related areas, and none of them have spent a critical period outside of Korea. Additionally, as a non-expert evaluator group, 8 Korean students with high-level English proficiency were selected to compare with the professional evaluator groups (ER and KR). All KG raters consisted of students with high English proficiency with a TOEIC score of 950 or higher and less than one year of experience in English speaking countries.

2.2 Stimuli

2.2.1 Speaker for the assessment material

In order to eliminate the factors other than speaking rate that could affect the assessment results, including segmental features, intonation, and rhythm, etc., test tokens from a native English speaker were chosen for generating evaluation stimuli. Assuming that the pronunciation, intonation and rhythm of native speakers were perfect, the assessment would not be adversely affected like those of L2 speakers.' Speech data from one native English speaker were selected to generate evaluation stimuli. The speaker was in his mid-thirties and was born and raised in New York. He is a professional speaker with a lot of experience in recording for various English tests including official English tests. In addition, the limited range of speaking rates, which was due to computer modified manipulation, is solved by using natural human utterances. The evaluation samples were recorded using an AKG C520 microphone and TASCAM US-144MKII interface in 16kHz 16bit PCM.

2.2.2 Test materials

Since speaking rate is highly related to utterance length (Dankovičová 2001; Quené 2008; Jacewicz et al. 2010), in order to observe the evaluation trend of speaking rate according to the utterance length, three types of utterance lengths are applied as presented in Table 1.

Table 1. Test materials

Utterance length	Number of syllables	Test material
Short	7	John saw his lanterns last night.
Mid	15	Joseph the great was playing with his lanterns last Wednesday morning.
Long	27	This awful show Lina has gone over this year was "Sane and City in Libya" which was hidden for one year.

As most of the relevant previous studies employed utterances with 7 to 30 syllable lengths (Munro and Derwing 1998; Zellner 1998), this study also tried to match that. A syllable length of 27 was set for the maximum length, not only because it is difficult to maintain the same speed equally across the entire utterance (Brindley and Slatyer 2002), but also because there is the opportunity to insert pauses naturally due to the long utterance length.

To achieve the goal of producing assessment stimuli as close to the natural sound as possible with minimum manipulation, first, the native English speaker was instructed to produce each sentence in Table 1 naturally. Then he was requested to produce faster speech samples for each sentence to several levels incrementally. In the same way, slower speech samples of varying degrees were also produced. This procedure was carried out three times in a row. The native speaker produced various speed levels from the fastest speaking rate of 8.66 syl/sec to the slowest, 1.56 syl/sec. Among these various levels of speech rates, more than half of the utterances were removed by excluding the similar rates of speed, and only 7 speech rates were chosen for each utterance length such that the pairwise speech rate interval between adjacent samples increased by approximately 120%, respectively. Table 2 presents the 7 levels of raw articulation rate (i.e. the speech rate which measured without pauses) for the

test.

Table 2. Raw articulation rate for test stimuli (syl/sec): r1 (the fastest) to r7 (the slowest)

Utterance length	r1	r2	r3	r4	r5	r6	r7
Short (7 syl)	6.58	5.55	4.50	3.54	2.87	2.05	1.56
Mid (15 syl)	7.20	6.52	5.45	4.29	3.22	2.64	2.37
Long (27 syl)	8.66	7.74	5.83	4.80	4.03	2.98	2.28

As shown in Table 2, the speed at which the speaker could speak differed depending on the length of the utterance. This may be due to different limits that can be condensed or increased for each phoneme, or to the widely-known general tendency of native speakers to alter speaking speed according to the length of the sentence. Longer phrases tend to have faster articulation rates as speakers shorten the length of syllables to produce more syllables within a phrase (Nooteboom 1972; Lindblom and Rapp 1973; De Rooij 1979; Nakatani et al. 1981, as cited in Quené 2008).

In Table 2, 'r1' indicates the fastest speech rate for each utterance length. Compared to r1, the speech rate is approximately 1.25 times slower when it goes to next level 'r2.' Likewise, from r2 to r3, and r3 to r4, etc., it slows down approximately 1.25 times at each interval. Since it is normally very odd to speak at less than 1 syl/sec, and almost impracticable to speak faster than 8 syl/sec, it can be said that almost every stage of speech rate that humans can produce are included on the test materials. In this way, the various speed ranges of natural-sounding test tokens could be obtained without distortion of sound caused by computer manipulation.

In summary, all 21 utterance tokens (3 utterance lengths × 7 speeds = 21) are employed for the test together with other dummy samples.

2.3 Manipulation : Adjusting speech rate

As pausing in slow speed is a natural phenomenon, a natural pause was presented in the raw tokens, especially in slow speed ranges.[2] However, since

2 It is widely known that pauses are used as a strategy for decreasing speech rate (Goldman-Eisler 1968; Griffiths 1990; Zellner 1994).

the evaluation results can be biased by the pause appearance itself, first of all, the pauses that appeared in the original utterance were eliminated. An additional handling was applied to the raw stimuli in order to remove the influence of prepausal lengthening as much as possible. The lengthened syllables derived from pause appearance needed to be shortened by as much of their lengthened ratio, which is widely known as 'prepausal lengthening.' Prepausal lengthening is a well-known durational effect by which pauses influence the length of the preceding units (Klatt 1976; Crystal and House 1988; Campbell 1991).

However, it is beyond the scope of this study to obtain the exact prepausal lengthening ratio at each specific speed (all 21 conditions, 3 utterance lengths × 7 speed conditions), with even different segmental values. For this reason, in order to offset the effects of the prepausal lengthening as much as possible, the relevant speaker's own mean value of lengthened ratio is employed for adjusting the lengthened duration.

To calculate the prepausal lengthening ratio, the speaker was instructed to read sentences with several pauses, then the same sentences without pauses. By having him read several times for each of the three speed conditions (slow, normal, and fast, see Table 3), the prepausal lengthening ratio was obtained by comparing the same words in prepausal positions with those in non-prepausal positions.

Table 3. The speaker's mean prepausal lengthening ratio of three different speech rate conditions

Articulation rate (syl/sec)	**Slow (3.34)**	**Nor (4.62)**	**Fast (6.30)**	**Mean**
Prepausal lengthening ratio	1.18	1.22	1.38	1.26

Table 3 indicates, in slow speech, for example, when the speaking rate is 3.34 syl/sec, that the words in prepausal positions are lengthened approximately 1.18 times longer than the words in non-prepausal positions. Based on this, the lengthening magnitude at various speed conditions can be estimated as shown in Figure 1.

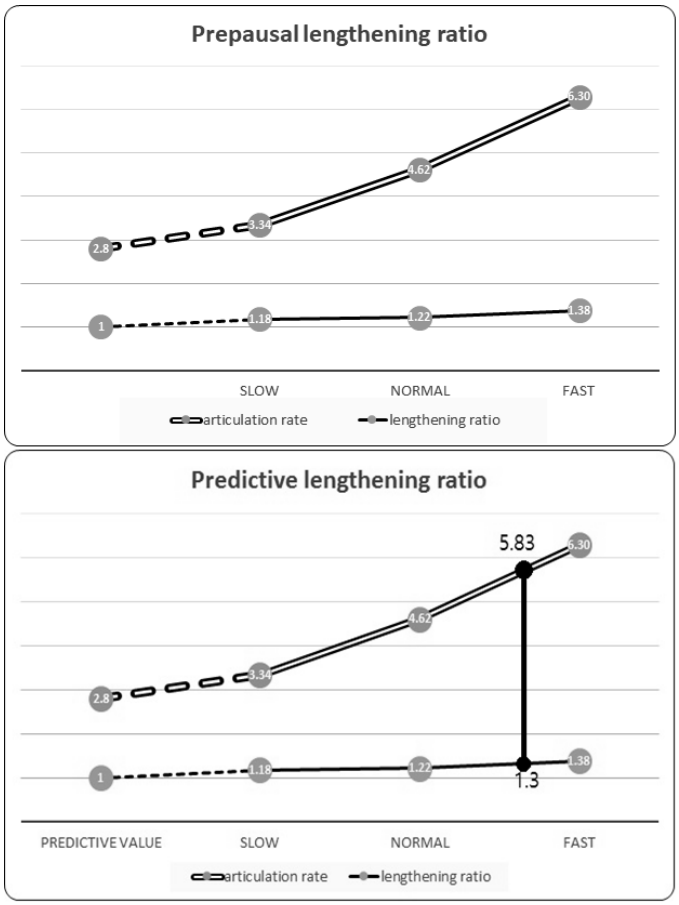


Figure 1. The ratio of prepausal lengthening at three articulation rate conditions vs. predictive lengthening ratio

For example, we can estimate the lengthening ratio for speaking rate of 5.83 syl/sec (r3-long utterance, see Table 2) as illustrated in Figure 1. That is, at speeds of 5.83 syl/sec, the word in prepausal position is elongated by approximately 1.3 times. This ratio is employed for reverting the lengthened syllable to its original length as similarly as possible. Table 4 shows how to offset the lengthened syllable by its duration before it lengthened.

As the pauses partly appeared only in slow utterances, and at very slow speech rate, pauses had no particular effect on preceding syllables; thus, only the

values of the italicized targets in Table 5 are manipulated.

Table 4. The samples of shortened syllable durations at prepausal positions of articulation rate stage "*r3* (5.83 syl/s)"

r3 utterance (5.83 syl/s)	**Raw syllable duration (sec) - lengthened by pauses**	**Shortening (raw syllable duration / 1.3)**
this	0.26	0.26
year	***0.22***	***0.17***
<pause>		
Li	0.18	0.18
bya	***0.30***	***0.23***
<pause>		

According to our relevant speaker's data (Figure 1), if the speaking rate is below approximately 2.88 syl/sec, the prepausal lengthening effect disappears. It appears to be that the phonemes were elongated to their maximum duration, with or without pause, due to the slow speed. For this reason, if the speed is slower than 2.88 syl/sec, only the pause was excluded.

The modified length is allocated in a syllable-based unit using Praat manipulation script. Since lengthening affects the final syllable the most (Shattuck-Hufnagel and Turk 1998), only prepausal syllables were targeted for modification. Looking at the italicized figures (which are modified targets) in Table 5, the final articulation rate after duration modification ('Predictive articulation rate' in Table 5) is not significantly different from the original speed ('Raw articulation rate' in Table 5).

Table 5. The predictive articulation rate (syl/sec)

Raw articulation rate							
Utterance length	**r1**	**r2**	**r3**	**r4**	**r5**	**r6**	**r7**
Short (7 syl)	6.58	5.55	4.50	3.54	2.87	2.05	1.56
Mid (15 syl)	7.20	6.52	5.45	4.29	3.22	2.64	2.37
Long (27 syl)	8.66	7.74	5.83	4.80	4.03	2.98	2.28
Predictive articulation rate							
Short (7 syl)	6.58	5.55	4.50	3.54	***2.88***	2.05	1.56
Mid (15 syl)	7.20	6.52	5.45	4.29	***3.30***	2.64	2.37
Long (27 syl)	8.66	7.74	***5.99***	***5.01***	***4.18***	***3.06***	2.28

As the shortening ratio was relatively small, the manipulated speech tokens were very natural sounding. Other than this, no other artificial means of controlling speed were applied, so many more natural-sounding materials could be obtained compared to those used in previous relevant studies.

2.4. Procedure

The evaluators were given simple printed evaluation criteria that referred to various scoring guidelines of speaking tests (TOEIC and FLEX, etc.) and Riggenbach (1991), and the three sentences to be evaluated (See Table 1) were also given in print (See Appendix A). Evaluators were asked to regard the experiment as an actual assessment. A ten-point rating scale was employed as normally a 9 to 12-point Likert scale is adopted in international speaking tests. Using Praat Experiment MFC, evaluators clicked the score button on a computer directly (See Appendix B for the screenshot of experiment).

The pretest was conducted prior to the regular evaluation, and test raters could replay the target sounds and also could re-rate previous utterances.

However, the KG Group was divided into two groups, each conducting an experiment at the same time on their own answer sheets, so if someone wanted to replay the test utterance, the other evaluators also had to listen to the same test utterance again. The experiment, including dummy samples, took about 13 minutes for most evaluators.

3. Results and discussion

The data analysis software, *R* (version 3.2.5) was used for statistical analysis. The linear mixed-effect model fit through R library package *lsmeans* (Lenth 2015).

3.1 Speech rate

As presented in Figure 2, the three groups presented a similar evaluation pattern. They assigned better scores for tokens with the rate between 4 and 5 syl/sec, which reflects native speakers' average speech rate. Moreover, evaluators

scored low on not only very slow speed but also very fast speed. However, among the group of experts, faster speech was assigned somewhat better points than slower speech.

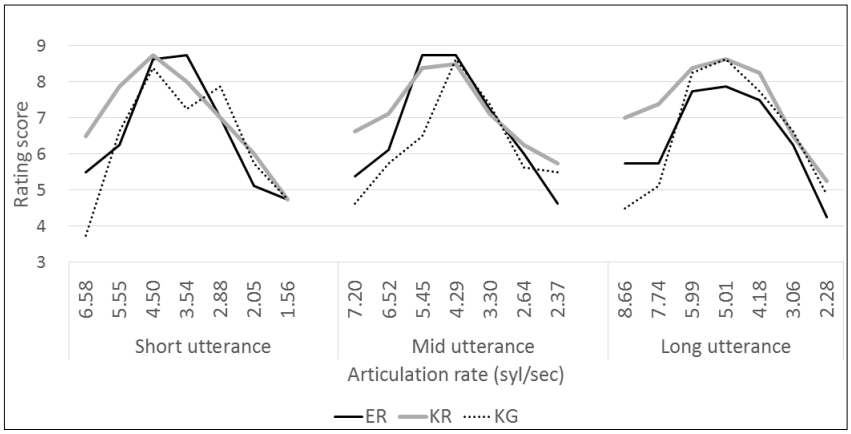


Figure 2. Mean evaluation scores for speech rate

As shown in Figure 2, the KR group showed a tendency to give higher scores at a faster rate than the other groups, but the KG group, conversely, tended to give lower scores to faster speeds than slower ones. Though the ER group tended to assign lower scores to slower speeds compared to other groups, their evaluation of slow speed ranges showed a similar tendency to the other groups; statistically, there were no meaningful differences found between the three groups. This seems to be because all three groups have high English proficiency.

To assess the reliability of the evaluator group, 'ICC (Intraclass Correlation Coefficient)' was adopted using *R* (version 3.5.3) and its *irr* (Shrout and Fleiss 1979) package. The inter-rater reliability of the non-expert student group (0.66, $F(20, 140) = 16.6$, $p<0.0001$) was higher than that of the expert group (0.61 for ER, $F(20, 140) = 13.5$, $p<0.0001$; 0.43 for KR, $F(20, 140) = 7.13$, $p<0.0001$)

As the number of evaluators was small and as there were also some areas given discrepant scores among raters in the same group, especially the KR group, the overall reliability seems to be low (See 3.3 for the discrepant

evaluation results of the KR group).

The range of speech rate preferences of the three groups are presented in Table 6.

Table 6. The range of speech rate preference (syl/sec)

Score	Mean	ER	KR	KG
8 or above	3.54 - 5.99	3.54 - 5.45	3.54 - 5.99	4.50 - 5.99
7 or above	2.88 - 5.99	2.88 - 5.99	2.88 - 8.66	2.88 - 5.99
6 or above	2.88 - 7.74	2.64 - 6.52	2.05 - 8.66	2.88 - 5.99

It can be seen that it is difficult to get a high score with a speech rate that is very fast or slow, outside the normal speed range. The group of professional evaluators assigned high scores to a wider range of speeds than the non-expert group. That is, if speakers do not deviate significantly from the normal speed range, professional raters do not penalize them much. The results indicate that it is difficult to expect high scores when an utterance is slower than 2.8 syl/sec or faster than 6 syl/sec.

Based on these results, L2 learners who are preparing for English speaking tests, do not have to be too pressured by the speed of speaking unless they are too slow, given the typical L2 speakers' speaking speed (around 3 to 4.5 syl/sec)[3].

As shown in Table 7, the highest score given by the ER group was 8.75 on average, which is 3.54 syl/sec (short utterance), 4.29 and 5.45 syl/sec (mid utterance; in long utterance, 5.01 syl/sec got the highest score but only scored 7.88 on average). The speaking rate that drew the best score for KR was 4.50 syl/sec (mean score 8.75, short utterance), and for KG it was 4.29 syl/sec (mean score 8.63, mid utterance) and 5.01 syl/sec (long utterance). In terms of averages, the speaking rate that drew the highest score for ER was 4.43 syl/sec, with 4.50 syl/sec for the KR group, and 4.65 syl/sec for the KG group.

3 The average speech rate for a non-native speaker is around 3 to 4.5 syl/sec [4.4 syl/sec for high level L2 speakers, 3.3 for intermediate level in Anderson-Hsieh and Venkatagiri 1994, 3.8 in Munro and Derwing 1998, and approximately 3.59 syl/s (160 wpm) in Götz 2013].

Table 7. Average score by speech rate in three groups

	Speech rate (syl/sec)	ER	KR	KG
Short utterance	6.58	5.50	6.50	3.75
	5.55	6.25	7.88	6.63
	4.50	8.63	8.75	8.38
	3.54	8.75	8.00	7.25
	2.88	7.00	7.00	7.88
	2.05	5.13	6.00	5.75
	1.56	4.75	4.75	4.75
Mid utterance	7.20	5.38	6.63	4.63
	6.52	6.13	7.13	5.75
	5.45	8.75	8.38	6.50
	4.29	8.75	8.50	8.63
	3.30	7.25	7.13	7.38
	2.64	6.00	6.25	5.63
	2.37	4.63	5.75	5.50
Long utterance	8.66	5.75	7.00	4.50
	7.74	5.75	7.38	5.13
	5.99	7.75	8.38	8.25
	5.01	7.88	8.63	8.63
	4.18	7.50	8.25	7.75
	3.06	6.25	6.50	6.63
	2.28	4.25	5.25	4.88

Table 8 shows the post hoc test results for speech rate. Looking at the first column of the first table in Table 8, the mark '*' indicates that the difference in evaluation scores between r1 and r2 in a short utterance is significant to the p value below 0.1. As can be seen, for an appropriate speaking rate range (r3 ~ r4 and r4 ~ r5), evaluators do not differentiate their evaluation scores statistically. The statistically meaningful results were mainly evident at boundaries of very slow or very fast speed when the difference in speed was at least over 1 syl/sec. Normally, a gap of less than 1 syl/sec showed no statistically significant output. As reported in Kendall (2013), listeners perceive differences in speaking rate ± 0.25 syl/sec, so it seems that raters do not strongly take into account a difference of less than 1 syl/sec in their judgement if the utterance speed falls within the appropriate speed range.

Table 8. The post hoc test results for speech rate:
'NS' denotes 'Not Significant'
'***' p<0.0001, '**' p<0.001, '*' p<0.01, '.' p<0.05

Short $x^2(6)=97.103$, $p<0.0001$

Short	**r2**	**r3**	**r4**	**r5**	**r6**	**r7**
syl/sec	5.55	4.5	3.54	2.88	2.05	1.56
r1 / 6.58	*	***	***	**	NS	NS
r2 / 5.55		*	NS	NS	NS	***
r3 / 4.50			NS	NS	***	***
r4 / 3.54				NS	***	***
r5 / 2.88					*	***
r6 / 2.05						NS

Mid $x^2(6)=78.389$, $p<0.0001$

Mid	**r2**	**r3**	**r4**	**r5**	**r6**	**r7**
syl/sec	6.52	5.45	4.29	3.3	2.64	2.37
r1 / 7.20	NS	***	***	*	NS	NS
r2 / 6.52		*	***	NS	NS	NS
r3 / 5.45			NS	NS	**	***
r4 / 4.29				.	***	***
r5 / 3.30					.	**
r6 / 2.64						NS

Long $x^2(6)=95.385$, $p<0.0001$

Long	**r2**	**r3**	**r4**	**r5**	**r6**	**r7**
syl/sec	7.74	5.99	5.01	4.18	3.06	2.28
r1 / 8.66	NS	***	***	***	NS	NS
r2 / 7.74		***	***	**	NS	.
r3 / 5.99			NS	NS	*	***
r4 / 5.01				NS	***	***
r5 / 4.18					.	***
r6 / 3.06						*

3.2 Utterance length

Though statistically meaningful results were not reported in utterance length, the longer the utterance length, the higher the score tended to be. Also, evaluators tend to be more lenient to slow speeds in short utterances and faster speeds in long utterances. This seems to reflect the tendency of native speakers, as they tend to speak longer utterances at a faster pace to produce more

syllables within a limited amount of time (Quené 2008). This is because speech that is too slow is difficult for the listener to process (Munro and Derwing 1998). A proper semantic unit should be understood at once by listeners for the sake of semantic transmission; however, speech that is too slow interferes with this process, making it difficult for the listener to understand what is being said. Since speaking long utterances at a fast pace is a natural phenomenon from a cognitive point of view, this seems to have been reflected in the assessment results.

3.3 Differences in evaluation tendencies among groups

It is noteworthy that there were rather inconsistent rating results at both a very fast and a very slow speech rate, compared to other speed ranges, which could lead to discrepant rating results among raters. Figure 3 presents the interesting results of evaluators' distinctive tendency at speeds that were too fast or too slow. For example, evaluator SJJ, one of the KR evaluators, assigned the highest score at extremely fast speeds, while YJH, also a KR evaluator, gave the lowest score to them.

Table 9 shows the number of evaluators for different speed preferences. As can be seen, expert raters preferred faster speeds more, while non-expert raters preferred slower speeds. In Table 9, the meaning of '2' below the 'Fast' in the first row indicates that in the ER group, two evaluators gave higher scores to faster speeds over slower ones (like the data of 'KR_SJJ' in Figure 3). The '6' indicates that six evaluators showed similar patterns to those of 'KR_LJH' in Figure 3.

Table 9. The number of raters' preference difference

	Fast	Equal variance	Slow
ER	2	6	0
KR	5	1	2
KG	1	2	5

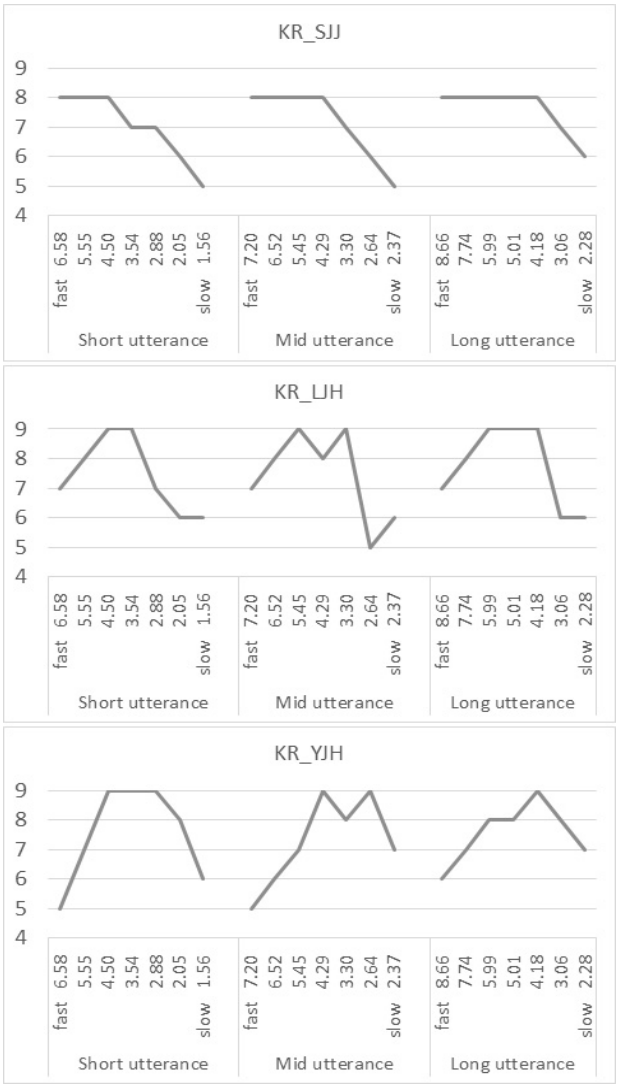


Figure 3. The scores for speech rate showing partiality of individual raters from the KR group

What is unusual is that there was no native-speaker evaluator who gave a higher score to a slow speed over a fast speed. Native speakers had the most equal variance, which indicates that native speakers are more concerned with the

proper speed range, and Koreans evaluators do not consider the fast speed to be a point of deduction even if it goes beyond the normal speed category. This strong disfavor of extremely fast or slow utterance speeds leads to inconsistent rating outcomes among raters. Though most expert rater groups preferred faster speeds more, this issue should be discussed in the assessment rubric to avoid discrepant rating results.

Moreover, since there was a difference between groups of evaluators at extremely high speeds, L2 learners do not necessarily have to read or speak at very high speeds to get high scores. Considering the results of using the data from a native speaker with good segmental features, a very fast speed can be more perilous to non-native English speakers with problematic pronunciation. In fact, in my own experience, there are some cases in which students who speak very fast on English speaking tests to the detriment of other important evaluation factors, such as pronunciation, intonation, and rhythm.

4. Conclusion

This study aims to help L2 learners prepare for English speaking tests by finding the speaking rate of English utterances to which professional raters give the best scores. The experiment was conducted on professional evaluators with experience in official English test assessment to see how they evaluate various speaking speeds and how they differ from non-professional raters. The study also used closely controlled test stimuli to solve the problems found in previous studies of relevance, such as influence of awkward segmental features, rhythm, or intonation in the assessment resulting from using L2 speech data, and distortion and limited assessment speed range from using artificial speech control.

The speaking rate that drew the highest score by the expert group averaged 4.47 syl/sec [average 4.43 for an expert evaluator group of native English speakers (ER), and 4.50 for an expert evaluator group of Koreans (KR)], and by the non-professional group(KG), 4.65 syl/sec, which is a little bit lower than the preferred speaking rate reported in previous studies (4.7, 4.5, 4.7, 4.84, 4.69 syl/sec, see section 1).

Also, the results show that it is difficult to expect to get high scores when an utterance is slower than 2.8 syl/sec or faster than approximately 6 syl/sec. The most notable difference between groups is that the KR group gave a relatively high score for abnormally fast speech, whereas conversely, the KG group gave even lower points to the fastest speech than to the slowest speech. The ER group gave the lowest scores for slow speeds; however, compared to the other groups, they made a similar reduction for faster or slower speeds that were outside the appropriate speed range. This indicates that L2 learners do not necessarily have to read or speak at very high speeds to get high scores. Also, in order to reduce the discrepancies in rating scores based on speech rate, a more detailed and specific rubric for evaluators is necessary.

Although some emphasize the efficacy of automatic assessment (Neumeyer et al. 1996; Cucchiarini et al. 2000; Franco et al. 2000), there has been no consensus on the weight of the rating to be given to speech rate. An investigation employing quantitative analysis of a more varied range of speech rates would be beneficial for both human raters as well as for automatic assessment for determination of score weighting.

No matter how perfect a native speaker's utterance may be, it is hard to say that only the speech rate of the utterance was accurately assessed because it is almost impossible to remove all parameters affecting the rating results such as intonation, rhythm and pauses. However, when the speech gets faster or slower, many of these things change along with it. If they are all mechanically fixed, then naturally there will be additional awkward stimuli, making it impossible to discretely evaluate various speed ranges.

Moreover, test tokens using carefully controlled L2 speakers' utterances would give meaningful parallels to this experiment. It would be good to compare the equivalent test tokens using an L2 speaker's utterance with the current research to analyze the characteristics of L2 learners' speech that influence evaluation.

References

Anderson-Hsieh, Janet and Horabail Venkatagiri. 1994. Syllable duration and pausing in

the speech of Chinese ESL speakers. *TESOL Quarterly* 28(4): 807-812.

Apple, William, Lynn A. Streeter, and Robert M. Krauss. 1979. Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology* 37(5): 715-727.

Brindley, Geoff and Helen Slatyer. 2002. Exploring task difficulty in ESL listening assessment. *Language Testing* 19(4): 369-394.

Campbell, W. Nick. 1991. Durational shortening and anaphoric reference. In *Proceedings of International Congress of Phonetic Sciences (ICPhS)* 2: 286-289.

Crystal, H. Thomas and Arthur S. House. 1988. Segmental durations in connected-speech signals: Current results. *The Journal of the Acoustical Society of America* 83(4): 1553-1573.

Cucchiarini, Catia, Joost Van Doremalen, and Helmer Strik. 2010. Fluency in non-native read and spontaneous speech. In *Proceedings of Disfluency in Spontaneous Speech (DiSS) and Linguistic Patterns in Spontaneous Speech (LPSS) Joint Workshop,* Tokyo, Japan.

Dankovičová, Jana. 2001. *The linguistic basis of articulation rate variation in Czech*. Frankurt am Main: Hector.

De Rooij, Johan. 1979. *Speech punctuation: An acoustic and perceptual study of some aspects of speech prosody in Dutch.* PhD Dissertation. Rijksuniversiteit Utrecht.

Derwing, Tracey M., Murray J. Munro, Ronald I. Thomson, and Marian J. Rossiter. 2009. The relationship between L1 fluency and L2 fluency development. *Studies in Second Language Acquisition* 31(4): 533-557.

Derwing, Tracey M., Marian J. Rossiter, Murray J. Munro, and Ronald I. Thomson. 2004. Second language fluency: Judgments on different tasks. *Language Learning* 54(4): 655-679.

Foulke, Emerson and Thomas G. Sticht. 1966. Listening rate preferences of college students for literary material of moderate difficulty. *Journal of Auditory Research* 6: 65-89.

Franco, Horacio, Leonardo Neumeyer, Vassilios Digalakis, and Orith Ronen. 2000. Combination of machine scores for automatic grading of pronunciation quality. *Speech Communication* 30(2): 121-130.

Goldman-Eisler, Frieda. 1968. *Psycholinguistics: Experiments in spontaneous speech.* London: Academic Press.

Götz, Sandra. 2013. *Fluency in native and nonnative English speech* (vol. 53). Philadelphia/Amsterdam: John Benjamins Publishing.

Griffiths, Roger. 1990. Speech rate and NNS comprehension: A preliminary study in time-benefit analysis. *Language Learning* 40(3): 311–336.

Isaacs, Talia. 2008. Towards defining a valid assessment criterion of pronunciation proficiency in non-native English-speaking graduate students. *Canadian Modern Language Review/La Revue Canadienne des Langues Vivantes* 64(4): 555-580.

Jacewicz, Ewa, Robert A. Fox, and Lai Wei. 2010. Between-speaker and within-speaker variation in speech tempo of American English. *The Journal of the Acoustical Society of America* 128(2): 839-850.

Kang, Okim. 2010. Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System* 38(2): 301-315.

Kang, Seok-Han and Seok-Chae Rhee. 2011. A study on the Suprasegmental Parameters Exerting an Effect on the Judgment of Goodness or Badness on Korean-spoken English. *Journal of the Korean society of speech sciences* 3(2): 3-10.

Kendall, Tyler. 2013. *Speech rate, pause and sociolinguistic variation: studies in corpus sociophonetics*. New York: Palgrave Macmillan.

Kim, Mi-Sun and Tae-Yeoub Jang. 2016. The speaking rate of English that evaluators prefer. In *Proceedings of Korean Society of Speech Sciences (KSSS) 2016 Fall Conference*, 163-164.

Klatt, Dennis H. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *The Journal of the Acoustical Society of America* 59(5): 1208-1221.

Lass, Norman J. and Herbert A. Leeper. 1977. Listening rate preference: Comparison of two time alteration techniques. *Perceptual and Motor Skills* 44(3): 1163-1168.

Lee, Joo-Kyeong. 2018. Prosodic remedies of Korean talkers' English L2 speech: Optimal pitch and speech rate. *Linguistic Research* 35(1): 203-232.

Lehtonen, Jaakko. 1979. Speech rate and pauses in the English of Finns, Swedish-speaking Finns, and Swedes. In Rolf Palmberg (ed.), *Perception and Production of English: Papers on Interlanguage*, 35-51. Åbo: Åbo Akademi.

Lindblom, Björn E. and Karin Rapp. 1973. Some temporal regularities of spoken Swedish. In G. Fant and M. A. A. Tatham (eds.), *Auditory Analysis and Speech Perception,* 387-396. London: Academic Press.

Munro, Murray J. and Tracey M. Derwing. 1998. The Effects of Speaking Rate on Listener Evaluations of Native and Foreign-Accented Speech. *Language Learning* 48(2): 159-182.

Munro, Murray J. and Tracey M. Derwing. 2001. Modeling perceptions of the accentedness and comprehensibility of L2 speech the role of speaking rate. *Studies in Second Language Acquisition* 23(4): 451-468.

Nakatani, Lloyd H., Kathleen D. O'Connor, and Carletta H. Aston. 1981. Prosodic aspects of American English speech rhythm. *Phonetica* 38: 84–105.

Neumeyer, Leonardo, Horacio Franco, Mitchel Weintraub, and Patti Price. 1996. Automatic text-independent pronunciation scoring of foreign language student speech. In *Proceedings of International Conference on Spoken Language Processing (ICSLP),* 1457-1460.

Nooteboom, Sibout G. 1972. *Production and perception of vowel duration: A study of durational properties of vowels in Dutch.* PhD Dissertation. Rijksuniversiteit Utrecht.

Quené, Hugo. 2008. Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo. *The Journal of the Acoustical Society of America* 123: 1104-1113.

Riggenbach, Heidi. 1991. Toward an understanding of fluency: A microanalysis of nonnative speaker conversations. *Discourse Processes* 14(4): 423-441.

Shattuck-Hufnagel, Stefanie and Alice Turk. 1998. The domain of phrase-final lengthening in English. In *The Sound of the Future: A Global View of Acoustics in the 21st Century, Proceedings of the 16th International Congress on Acoustics and 135th Meeting Acoustical Society of America*, 1235-1236.

Shrout, Patrick E. and Joseph L. Fleiss. 1979. Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin* 86(2): 420-428.

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

Zellner, Brigitte. 1994. Pauses and the temporal structure of speech. In Eric Keller (ed.), *Fundamentals of speech synthesis and speech recognition,* 41-62. Chichester: John Wiley.

Zellner, Brigitte. 1998. Temporal structures for fast and slow speech rate. In *The Third ESCA/COCOSDA Workshop (ETRW) on Speech Synthesis,* 143-146.

Appendix A. Printed evaluation criteria

<Evaluation Criteria>

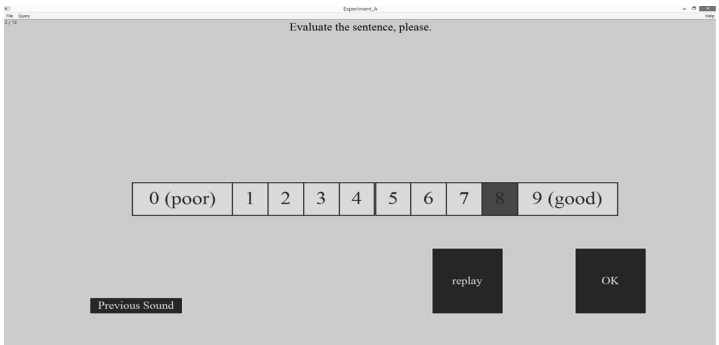
-Pronunciation (audible sounds, good word linking, no mistakes, no influence of other language,)
-Fluency (affective intonation and stress, no frequent short or long pauses, no hesitation, no repair, appropriate speech rate)

Evaluation Criteria	Pronunciation fluency	Hardly audible Bad intonation and stress Frequent pauses								Distinctive and easily audible Good word linking Affective intonation and stress No frequent pauses
	point	0 (very poor)	1	2	3	4	5	6	7	8	9 (very good)

* Read a text aloud.

1. John saw his lanterns last night.
2. Joseph the great was playing with his lanterns last Wednesday morning.
3. This awful show Lina has gone over this year was 'Sane and City in Libya' which was hidden for one year.

Appendix B. Screenshot of experiment



Mi-Sun Kim
Instructor
Dept. of English Language and Language Technology
Hankuk University of Foreign Studies
107, Imun-ro, Dongdaemun-gu,
Seoul, 02450, Rep. of Korea
E-mail: misunk@hufs.ac.kr

Received: 2019. 05. 20.
Revised: 2019. 09. 09.
Accepted: 2019. 09. 12.