

Effects of L1 Raters' Language on Fluency Assessment for L2 Suprasegmentals

Seokhan Kang*
(Seoul National University)

Kang, Seokhan. (2012). Effects of L1 raters' language on fluency assessment for L2 suprasegmentals. *The Linguistic Association of Korea Journal*, 20(4), 25-44. This study investigated how the various rating groups with different mother tongues judge the fluency for the L2 speaking. It was expected that the Korean raters have different rating patterns from native English raters for Korean speaking of English. Fourteen raters took part in fifteen speaking samples. After L2 subjects' speech was assessed by two different rating groups, acoustic measurement of L2 speech was analyzed. The results show that the effect of raters' L1 background language exhibits differently in rating the L2 fluency. More specifically, Korean raters put more focus on pause duration in the fluency evaluation, while native English raters focused on speech rate. It means that their rating patterns could be different depending on the raters' background language.

Key Words: fluency, mother tongue, evaluation, speaking, prosody, F0 range, speech rate, pause duration

1. Introduction

The suprasegmental features of an L2 (a second language) are considered to be the major factor in judging the fluency of L2 speech. More specifically, L2 fluency might be influenced to a greater extent by suprasegmental factors (Anderson-Hsieh, Johnson, & Koehler, 1992; Munro & Derwing, 1995; Derwing, Munro, & Thomson, 2009). Also forming the L2 suprasegmentals is closely tied

* I really appreciate the chief-editor and the reviewers for their thoughtful comments.

with the variable of the subjects' background language (Derwing & Munro, 2001; Trofimovich & Baker, 2006). The aim of the study presented here is to investigate to what extent the L1 background language impacts on the L2 oral fluency. To examine the raters' L1 effect, this study analyzes the suprasegmental cues involved in the judgment.

Comparatively large studies investigating the acquisition of a second language have focused on what suprasegmental factors affect the L2 fluency. The factors that researchers have found to be significant in fluency assessment include speaking rate (Guion, Flege, James. Liu, Yeni-Komsian, 2000; Derwing & Munro, 2001), pause structure (Anderson-Hsieh & Vengatagiri, 1994; Riazantseva, 2001), peak alignment (Mennen, 2006; Trofimovich & Baker, 2006), F0 range (Wennerstrom, 1994; Kang, Guion-Anderson, Rhee, & Ahn, 2012), and intonation (Wennerstrom, 2002; Chun, 2002).

The L2 suprasegmentals, associated with the fluency judgment, are known to be affected by the effect of subjects' L1 background language. For instance, Purcell & Suter (1980) have examined the effect of L1 background language on the degree of fluent L2 speech. In the study, native speakers of Arabic and Persian were found to have a better pronunciation of English than native speakers of Japanese and Thai. This study provides evidence that transfer or interference in the L1 is an important factor in the production of L2 prosody. Wennerstrom (1994) investigated the English prosody produced by native Spanish, Japanese, and Thai speakers and reported that while the native speakers made significant use of pitch contrasts to signal focus on the items measured, the non-native speakers did not consistently use pitch to signal meaning contrasts in many of the same environments. The study implies that L2 prosody formation could be different according to the effect of L1 background.

Previous studies support the hypothesis that raters' L1 effect influences the evaluation of L2 speaking performance. It is assumed that the evaluation of L2 speaking is different depending on who evaluate the exams. The assumption has long been observed for native speakers as L2 speaking evaluators (Taylor, 2006; Kim Y-H, 2009). Naturally, large-scale and high-stakes tests such as TOEFL or OPICs rendered their assessments using native English speakers as a modeling benchmark (Lowenberg, 2000, 2002).

However, some researchers have raised questions on whether native English

raters should be the only acceptable standard just because they are native English speakers (Taylor, 2006). Indeed, it is hard to set up the clear definition of the native speakers of English. Furthermore, non-native English speakers outnumber native English speakers internationally (Crystal, 2003; Lowenberg, 2000), and variation of English has been extended in the world (Kachru, 1985, 1992). The environmental change calls for the non-native speakers' participation in rating non-native speakers' English proficiency tests.

The inevitable requirement for non-native raters' involvement in non-native speakers' proficiency raised some doubt on validity and reliability for their ratings (e.g., Kim J-K, 2006, Lee C-H, 2010, Kang & Rhee, 2011). However, previous works have been focused on how the Korean raters' rating patterns are different from those of native English raters in both tasks and criteria (e.g., Kim Y-H, 2009; Lee C-H, 2010, Kang & Rhee, 2011). Rather, this study used more in-depth analysis for the L2 suprasegmentals on how their rating results were influenced by the L1 background language. This study tried to suggest a better objective method in assessing the English speaking performance by Korean raters. The practical goal of the study is to analyze the Korean raters' assessment patterns comparing with native English raters on the fluency of L2 speech. Thus, this research would answer the following questions;

- (1) What distinctive evaluation patterns do Korean raters exhibit on the fluency judgment of an L2, compared with native English raters?
- (2) How do these L2 suprasegmentals affect the fluency judgment of two different rating groups?

2. L1 effect on L2 fluency judgment

Studies on how native and/or non-native raters approach the tasks of the fluency assessment for L2 speaking are still relatively few in Korea. Fluency is defined as the listeners' judgment of how natural an utterance sounds, spoken without undue pauses, filled pauses, hesitations, slow speech rate, or dysfluencies (Derwing & Munro, 2001; Trofimovich & Baker, 2006). Some researchers argue that L2 learners' mother tongue influences on the fluency of

the target language. For instance, Fayer and Krasinski (1987) examined how the English-speaking performance of non-native English students was perceived by native and non-native English raters. The results showed that non-native raters tended to be more severe in general and to express more annoyance when rating linguistic forms and that pronunciation was the most distracting factors for both sets of raters. Kang and Ahn (2012) compared Korean raters with native English raters for English speaking assessment of Koreans. The results of the experiments show that most of the raters in both groups maintained acceptable level of internal consistency, regardless of their L1 background. The non-native Korean raters show similar severity patterns across various tasks and criteria, along with native English raters. However, Korean raters exhibit some particular characteristics different from native English counterparts, in which English raters exhibited strict measurement on discourse cohesion, while Korean raters showed the lenient measurement in this task.

Most of the previous works that have investigated the rating patterns between native and non-native raters have been conducted using only either limited tasks or criteria. However, this study, examining the L2 fluency, tried to suggest that Korean speaking of English could be affected by raters' background language. For the proof, the L2 suprasegmentals have been analyzed in both production and perception.

3. Methodology

3.1. Participants

Fifteen Korean students were selected from a college-level English class of a university in Seoul, Korea, and were informed about the research project. The students (male 8, female 7) were assigned to mid-level classes based on the English proficiency test before beginning the regular classes of the semester. It means that each sample would be considered as almost equal level of English proficiency.

The test was administered in a computer-mediated indirect interview format. The indirect method was chosen because of its effectiveness, reliability, and easy

accessibility (e.g., Flege, 2006). The test lasted approximately 30 minutes. We paid each participant an honorarium of small money upon completion of the experiment. Following is the participants' information.

Table 1. Participants' information

ID	Gender	Age
K1	F	21
K2	M	20
K3	F	19
K4	F	21
K5	M	22
K6	F	22
K7	M	20
K8	F	19
K9	M	20
K10	M	19
K11	M	20
K12	M	20
K13	M	21
K14	F	21
K15	F	20

For the assessment, fourteen raters of both groups rated the speaker' performance individually, listening to their speaking. They were given criteria and references. Most of the Korean raters hold doctoral degree in the field of the English language and literature, and have experience in teaching English in the universities for 4 to 12 years, ranged from 34 to 45 years old. Most of the native English raters hold master degree, and have experience in teaching English for 4 to 7 years, ranged from 31 to 38 years old. All raters reported normal hearing. We paid each person an honorarium of small money upon completion of the experiment. Following is the raters' information.

Table 2. Raters' information

	ID	Gender	Age (year)	Teaching (year)
Native English raters	E1	M	31	4
	E2	M	35	7
	E3	F	30	5
	E4	M	32	7
	E5	F	33	6
	E6	F	35	7
	E7	F	40	7
Native Korean raters	K1	F	27	6
	K2	F	31	7
	K3	M	35	9
	K4	F	28	4
	K5	M	44	12
	K6	F	36	5
	K7	M	29	6

3.2. Criteria

Most of the current English speaking tests have been evaluated in communicative cohesion, grammatical accuracy, lexical choice, and fluent pronunciation (Luoma, 2004). For example, IELTS (International English Language Testing System) evaluates examinees' fluency and cohesion, lexical resources, grammatical range and accuracy, and pronunciation (Kim, Y-H, 2009). TSE in TOEFL or the speaking test in TOEIC also covers similar criteria: effective communication, competent performance, and native-like communication (Luoma, 2004).

In this study, the area of fluency was based on EAP specialist raters' identification which has been considered as an overall valid and reliable measurement of the examinee's proficiency ability (Brown, Iwashita, & McNamara, 2002). Following is the definition of fluency distributed to the raters.

Fluency: It was identified as one of suitable measures such as filled pauses, unfilled pauses, repair, total pausing time, speech rate, and mean length of run. The examinee can express him/herself spontaneously at length with a natural colloquial flow, avoiding or backtracking around any difficulty so smoothly that the native listener is hardly

aware of it.

For the criteria given, a five-point rating scale was developed for rating the test-takers' fluency performance as 0, 1, 2, 3, and 4. Rating scale plays a crucial role to obtain the reliability and validity for the assessment (Lane, 2008; Lee, Kim, & Lee, 2011). Usually 9 to 12 point Likert scale has been adopted in international speaking tests. However, some tests adopt 5 to 8 point rating scale because they should consider test-takers' characteristics and purposes (eg., G-TELP: 5 points, ICAO: 6 points, FCE: 6 points). Since the test used in the study was designed for the research which focuses on analyzing the difference between two groups over fluency assessment, 5-point Likert scale which covers comparatively large range of sampling population was adopted.

3.3. Speaking tasks

Speaking tasks are to achieve a particular goal or objective in a particular speaking situation (Backman and Palmer, 1996). This study focuses on open-ended method which measures the effective use of language knowledge and strategic competence (Bachman and Palmer, 1996). The purpose of the experiment was to compare two rating groups in assessing the overall oral communicative language performance spoken by Korean learners of English. In order to assess the L2 suprasegmentals of test-takers, the opinion task was used like the following directions:

The test taker presents his/her opinion for 2 or 3 minutes on the given topics. He expresses his clear position on the topic and provides the reasons or criteria. Following is the example used in the study:

In this part of the test, you will give your opinion about a specific topic. Be sure to say as much as you can in the time allowed. You will have 45 seconds to prepare. Then you will have 60 seconds to speak.

Are the single people qualified to adopt a child? Can they be good parents? Do you support it or not? Why? Use reasons and specific examples to support your answer.

3.4. Data analysis

Three analyses were performed. The first analysis examined the extent to which the learners were able to produce L2 suprasegmentals fluently, as measured by overall fluency ratings. In this analysis, the sentences spoken by the subjects were presented to fourteen raters for evaluation. Then, the judgments were compared across two different rating groups.

The second analysis examined the extent to which the learners were able to fluently produce specific suprasegmentals: pitch range, speech rate, and pause duration. The results of the acoustic measurements obtained were compared. The purpose of the experiment was to analyze to what extent each measurement has been affected over the fluency levels.

The final analysis extended the findings of the first two by using a multiple regression procedure to investigate how the learners' production of specific suprasegmentals contributed to two different raters' fluency judgments in the L2 speech.

4. Ratings of fluency

4.1. Ratings and Raters

Samples from L2 speech recordings ensured that the content was held relatively constant across speakers. The fifteen samples were randomized and presented to the raters using a loud speaker. A total of fourteen raters were recruited to evaluate the fluency of the L2 speakers using a 5-point scale (from 0 = no fluent speaking to 4 = extremely native-like fluent speaking). The experimental sentences were presented to two different groups fluency judgment.

4.2. Results

Figure 1 showed that the raters varied widely in their measures of severity. In this study, English raters tend to rate harshly (0.21 of mean logit), while Korean raters have a tendency to grade leniently (-0.01 of mean logit). To check

the inter-rater reliability for both groups, Pearson correlation was obtained as $r=0.412$ ($p < .001$). It means that each rating was reliably scored among raters. Also to examine the intra-rater consistency for both groups, Spearman correlation was obtained as $r=0.534$ ($p < .001$) which refers to the consistent rating over each sample.

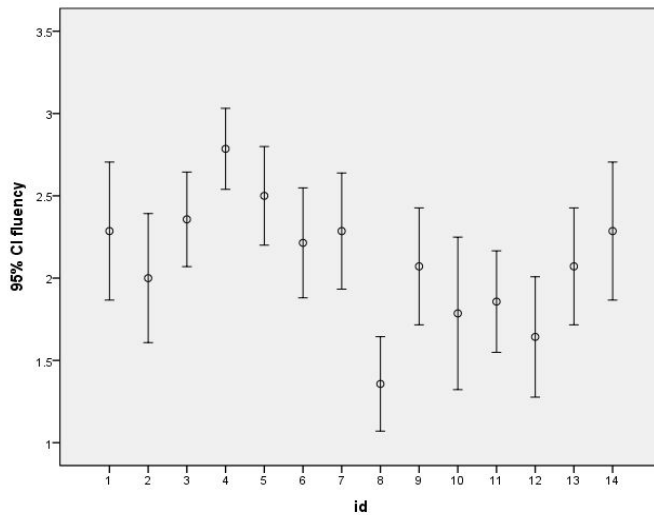


Figure 1. Fluency ratings for Korean raters (1–7) and English raters (8–14).

Korean raters grade somewhat leniently ranging from 1.75 to 3, whereas English raters grade severely ranging from 0.7 to 2.3 out of 4. The result was provided against the studies of Fayer and Krasinski (1987) and Santos (1989), in which non-native raters exhibit more strictness. Rather, it supports the results of Lee C-H (2010) and Kang and Ahn (2009) in that native English speakers exhibited more strict scores in judging the fluency and intelligibility. Native raters' harsh ratings may result from native raters' self-respect on the rating (Kim Y-H, 2009; Kang & Ahn 2012). On the contrary, the non-native raters may tend to raise the doubt to their own ratings, especially when performances are at the border of two adjacent fluency levels.

In short summary, the tests on raters' severity and internal consistency represented the particular Korean raters' characteristics: lenient measurement

and internal consistency. It means that Korean raters, along with native English raters, were consistent in their ratings, although they exhibited somewhat lenient ratings different from native English raters.

5. Acoustic analysis

The acoustic analysis sought to identify which specific suprasegmentals contributed to fluency ratings in the L2 learners' speech and to determine the extent to which they did so. Three specific suprasegmentals that might have contributed to the fluency judgment were examined in the acoustic analyses: F0 range, speech rate, and pause duration. The measured suprasegmentals represent the potential source contributing to L2 fluency ratings (Munro, 1995; Mennon, 2006; Trofimovich & Baker, 2006). In this study, measured acoustic cues were analyzed by following the perceptual category of good and bad response.

L2 utterance were used to evaluate the prosody of each group. Several acoustic measurements dealing with fundamental frequency (in Hertz) and duration (in milliseconds) were made. Duration and fundamental frequency were measured using a waveform display with a time-locked wideband spectrogram with the software praat (5.1.17). All acoustic cues were measured from the initial acoustic signal in both the waveform and the spectrogram to the final acoustic cues of the boundary such as burst or spectral cues (Kent & Read, 2003; Ladefoged, 2001).

5.1. F0 range

The F0 range is known as an indicator of English proficiency (e.g., Backman, 1979; Willems, 1982; Mennen, 2006). Generally, a lower proficiency of English as a second language is closely tied with a narrower F0 range. In this study, the range was measured from the highest point to the lowest point of the fundamental frequency: overall range of F0 across the intonational phrase. The analysis examined F0 range in the selected sentences to determine whether they were able to produce L2 F0 range natively with an increasing level of L2

proficiency and an L1 variation.

The F0 range obtained for each subject was submitted to a two-way ANOVA (L1 * evaluation). The analysis missed statistical L1 group effect, $p > .05$, as well as evaluation, $p > .05$. It means that two L1 rating groups showed no statistical difference in F0 range in fluency judgment. This result supports the proposal that F0 range is not one of the main judgment cues (e.g., Kang & Ahn, 2012; Kang, 2012). Figure 2 presents the mean range of F0 by two groups with different evaluations: bad or good. Note that the bar indicates the F0 range of bad or good pronunciation evaluated by both groups. Overall little difference could be found in both raters' classification of F0 range between good and bad pronunciation.

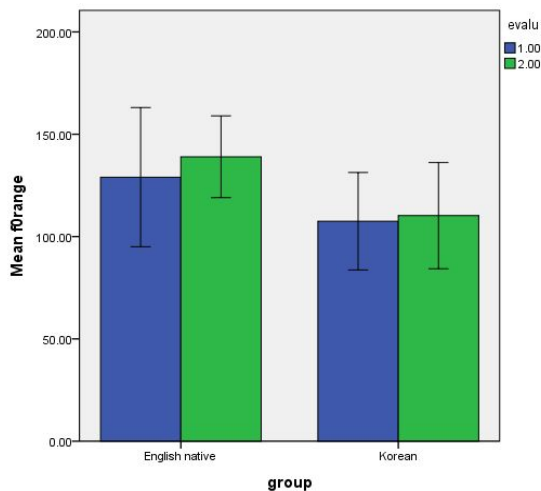


Figure 2. Group means of F0_range (± 1 SE) for both L1 groups' evaluation¹⁾.

5.2. Speech rate

Speech rate has been proved to be a good indicator of the second language proficiency (Derwing, Rossiter, & Thompson, 2004; Guion, 2005). Speech rate was calculated by dividing the total number of syllables produced in a given speech sample by the total time expressed in seconds (Ortega, 1999). In this

1) 1 refers to good evaluation, while 2 to bad evaluation.

study, the speech rate is operationalized as the duration measured from the initial acoustic signal of the phrase in both the waveform and the spectrograms to the final acoustic or spectral cues of the phrase boundary. Next, the resulting total number of syllables was divided by the total speech time excluding pauses of two or more seconds.

The two-way ANOVA revealed a significant group effect of L1 group, $F(1, 314) = 4.742$, $p < .05$, and evaluation, $F(1, 314) = 6.582$, $p < .05$. The interaction of L1 and evaluation was not significant as $F(2, 314) = 1.524$, $p > .05$. The Bonferroni tests exploring the simple main effect of L1 revealed that the native English raters grade good evaluation on the fast speech rate than on the slow speech rate ($p < .05$). The result agrees with previous works in which more native-like speech was produced with a faster speech rate (Adams & Munro, 1978; Lennon, 1990; Munro & Derwing, 1995; Sluijter & Van Heuven, 1996). Some research suggested that speech rate was not tied with L1 effect. For instance, Munro and Derwing (2006) reported that the listeners, regardless of L1 background language, had a tendency to assign the highest ratings to L2 speech that was somewhat faster than the rates generally used by L2 users, and also very fast and very slow speech tended to be less highly rated. In this study, however, Korean raters put less weight on speech rate in their judgement: little or no difference of speech rate between good and bad evaluation.

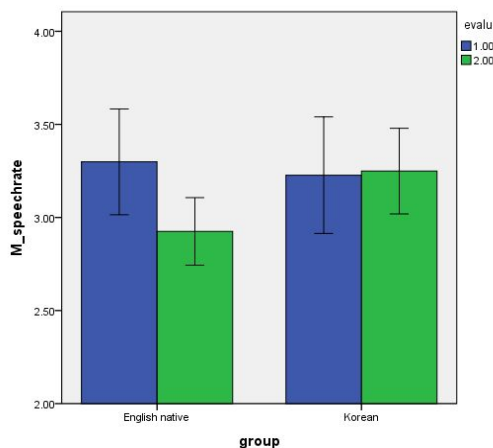


Figure 3. Group means of speech rate (± 1 SE) for both L1 groups' evaluation.

5.3. Pause duration

L2 acquisition studies report that both pause duration affect listeners' ratings of L2 fluency (Munro & Derwing, 2006; Rossiter, 2009), and that it is often viewed as a determinant of L2 fluency (Riazantseva, 2001; Trofimovich & Baker, 2006). Pauses might indicate the speaker's difficulty with the task, perhaps reflecting processing or memory constraint unique to L2 speech (Schachter, Christenfeld, Ravina, & Bilous, 1991; Riazantseva, 2001). It is suggested that pause duration is a cross-linguistic factor regardless of L1 background (Grosjean & Deschamps, 1975; Holmes, 1995). In this study, the pause duration as a total pausing time was calculated by adding up all the unfilled pauses, following the study of Iwashita, McNamara, and Elder (2001).

The two-way ANOVA revealed a significant group effect of L1 group, $F(1, 314) = 3.623$, $p < .05$, but it missed statistical evaluation effect, $p > .05$. The interaction of L1 and evaluation was not significant as $F(2, 314) = 0.982$, $p > .05$. The Bonferroni tests exploring the simple main effect of L1 revealed that Korean raters grade good evaluation on the short duration of pause than on the long duration ($p < .05$). The result agrees with previous works in which pause duration was one of the main cues to decide the L2 fluency (Riazantseva, 2001; Kang, Guion, Rhee, & Ahn, 2011). Figure 4 presents the mean range of pause duration by two groups with different evaluations: bad or good. Note that the bar indicates the pause duration of bad or good pronunciation evaluated by both groups. Overall the more they achieve the better fluency level, the shorter their pause duration range shows. However, the comparatively little difference could be found in English raters' classification of pause duration between good and bad pronunciation.

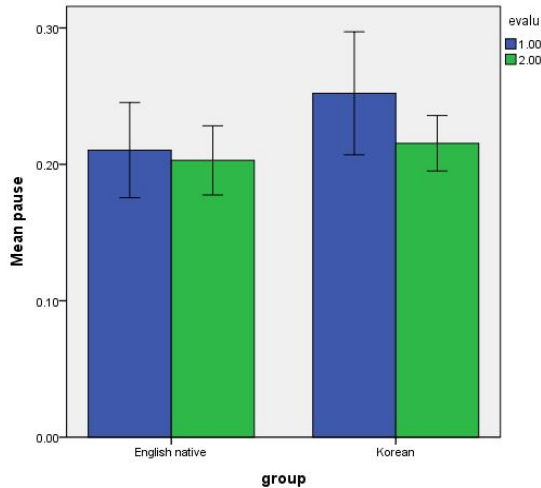


Figure 4. Group means of pause duration (± 1 SE) for both L1 groups' evaluation.

6. Relationship between suprasegmental production and fluency test

The production analysis reported that the effect of L1 influences L2 learners' judgment of fluency. One of the remaining questions on the production tests is to what extent their fluency for the suprasegmental cues could contribute to the judgment. For the analysis, both fluency rating scores of L2 learners of Korean and their values of suprasegmentals examined in this study were submitted to correlation and regression analyses. Zero-order correlations were computed between the subjects' fluency ratings ($n = 14$) and their suprasegmental measured values (F0 range, speech rate, and duration of pauses).

Table 3. Summary of correlation analyses between fluency ratings and acoustic measurement

		F0_range	Speech rate	Pause duration
Fluency ratings	English raters	.012	-.598**	-.136
	Korean raters	.009	-.178	-.412*

** $p < .001$, * $p < .05$

The analysis indicates that some of acoustic values measured in this study are differently correlated with the fluency ratings depending on raters' L1 background language (Table 3), suggesting that English raters put more weight on speech rate, while Korean raters on pause duration.

The fluency scores for each of the five suprasegmentals were individually regressed on the ratings for each L1 group. The goal of the analyses was to estimate the degree to which each of the three suprasegmentals examined here predicted the L2 learners' fluency ratings. These three separate linear regression analyses (Bonferroni adjusted $\alpha=.0125$) allowed for determining the amount of variance that each of suprasegmental measures shared with the fluency ratings.

Table 4. Summary of regression analyses for acoustic measurements in fluency ratings.

			B	SE B	B	R ²	T
English raters	Speech rate	Constant	3.672	.426			8.612**
		Variable	2.474	.270	.492	.378	4.572*
	F0range	Constant	26.274	6.723			3.284*
		Variable	10.001	4.570	.128	.016	.317
	Pause_ duration	Constant	.218	.055			3.995*
		Variable	.007	.034	.088	.176	.217
Korean raters	Speech rate	Constant	3.205	.492			6.517**
		Variable	.022	.311	.029	.166	.071
	F0range	Constant	37.750	13.570			2.470*
		Variable	2.750	6.192	.040	.165	.098
	Pause_ duration	Constant	.289	.068			4.616**
		Variable	2.382	.140	.654	.327	4.123*

** $p < .001$, * $p < .05$

Table 4 indicated that each of the three suprasegmental values, for English raters, significantly predicted the share of fluency ratings: the variable of speech rate explained 37.8 % of the variance and 17.6 % of the pause duration. On the contrary, the values of Korean raters predicted that the variable of pause duration explained 32.7% of the L2 fluency, along with each 16% of speech rate

and F0 range. It is clear that the temporal cues (e.g., speech rate, pause duration) are the most important in fluency judgment. However, cue weight is different depending on the raters' L1: speech rate for native English raters and pause duration for Korean raters. Also it remains to be questioned why some spectral signals such as the F0 range did not appear to have shared an important variance with L2 learners' fluency ratings. For the raters, fine-grained spectral differences in F0 range are so subtle that they are virtually imperceptible to them and then have little impact on fluency judgments.

In summary, the results suggest that native judgment of L2 speech may reflect the universal features of fluency judgment; a strong perceptual effect can exist on the temporal fluency such as the speech rate and the duration of the pause. However, which temporal cues are more influential is dependent on raters' background language.

7. Conclusion

This study has investigated the differences between both groups of native Korean and English raters in analytically judging performance on L2 suprasegmentals. It has revealed a little significant difference in the scores assigned by both groups. Although Korean raters were consistent in their ratings, they exhibited somewhat lenient rating different from native English raters.

The phonetic analysis exhibited some particular characteristics in cue weight dependency on L2 suprasegmentals. Some of acoustic values measured in this study are differently correlated with the fluency ratings depending on raters' L1 background language, suggesting that English raters put the more weight on speech rate, while Korean raters on pause duration. Overall, this study has improved our understanding on overall L2 speaking assessment by L2 raters. The study analyzes the current measurement as well as suggests the internal consistency, moderate severity on tasks, and reliable scaling division on the scoring for L2 speech.

References

- Adams, C., & Munro, R. (1978). In search of the acoustic correlates of stress: Fundamental frequency, amplitude, and duration in the connected utterances of some native and nonnative speakers of English. *Phonetica*, 35, 125-156.
- Anderson-Hsieh, J., Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable structure. *Language Learning*, 42, 529-555.
- Anderson-Hsieh, J., & Vengartagiri, H. (1994). Syllable duration and pausing in the speech of Chinese ESL speakers. *TESOL Quarterly*, 28, 807-812.
- Backman, N. E. (1979). Intonation errors in second language pronunciation of eight Spanish speaking adults learning English. *Interlanguage Studies Bulletin*, 4(2), 239-266.
- Bachman, L. F., & Palmer, A. F. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Bongaerts, T., Mennen, S., & Slik, F. (2000). Authenticity of pronunciation in naturalistic second language acquisition: The case of very advanced late learners of Dutch as a second language. *Studia Linguistica*, 54(2), 298-308.
- Brown, A., Iwashita, N., & McNamara, T. (2002). *Getting the balance right: Criteria in integrated speaking tasks*. Paper presented at the 24th Language Testing Research Colloquium, Hong Kong.
- Chun, D. M. (2002). *Discourse intonation in L2: From theory and research to practice*. Amsterdam, PA: John Benjamins Publishing Company.
- Derwing, T., & Munro, M. (2001). What speaking rates do nonnative listeners prefer? *Applied Linguistics*, 22(3), 324-337.
- Derwing, T., & Rossiter, M. (2003). The effects of pronunciation instruction on the accuracy, fluency, and complexity of L2 accented speech. *Applied Language Learning*, 13(1), 1-17.
- Derwing, T., Rossiter, M., & Thomson, R. (2004). L2 fluency: Judgments on different tasks. *Language Learning*, 54(4), 655-679.
- Derwing, T., Munro, M., & Thomson, R. (2009). A longitudinal study of ESL learners' fluency and comprehensibility development. *Applied Linguistics*, 29(3), 359-380.

- Flege, J. E. (2006). Phonetics approximation in second language acquisition. *Language Learning, 30*(1), 117-134.
- Gruttenden, A. (1997). *Intonation* (2nd ed.). New York: Cambridge University Press.
- Guion, S., Flege, J., James, E., Liu, H., & Yeni-Komshian, H. (2000). Age of learning effects on the duration of sentences produced in a second language. *Applied Psycholinguistics, 21*, 205-228.
- Guion, S. (2005). Knowledge of English word stress patterns in early and late Korean-English bilinguals. *Studies in Second Language Acquisition, 27*, 503-533.
- Iwashita, N., McNamara, T., & Elder, C. (2001). Can we predict task difficulty in an oral proficiency test? Exploring the potential of an information-processing approach to task design. *Language Learning, 51*(3), 401-436.
- Kachru, B. B. (1985). Standards, codification and sociolinguistic realism: The English language in the outer circle. In R. Quirk & H. Widdowson (Eds.), *English in the world: Teaching and learning the language and literatures* (pp. 11-30). Cambridge: Cambridge University Press.
- Kachuru, B. B. (1992). *The other tongue: English across cultures*. Urbana, IL: University of Illinois Press.
- Kang, S-H. (2012). A study on the fluency judgment of the suprasegmentals for the native English and Japanese learners of Korean. *Bilingual Research, 50*, 1-24.
- Kang, S-H., & Ahn, H. (2012). A comparative study on criteria and tasks in Korean English speaking assessment by native and non-native raters. *Language Research, 48*(2), 1-25.
- Kang, S-H., & Rhee, S-C. (2011). A study on the suprasegmental parameters exerting an effect on the judgment of goodness or badness on Korean spoken English. *Malsori and Speech Science, 3*(2), 3-10.
- Kang, S., Guion, S., Rhee, S., & Ahn, H. (2012). The effect of language immersion on the acquisition of second language suprasegmentals. *Applied Linguistics, 28*(1), 181-208.
- Kent, R., & Read, C. (2003). *Acoustic analysis of English*. Boston: Singular Thomson Learning.
- Kim, J-K. (2006). A validation study on the interview test in an university

- English program, *English Linguistics*, 6(1), 1-20.
- Kim, Y-H. (2009). An investigation into native and non-native teachers' judgements of oral English performance: A mixed methods approach. *Language Teaching*, 26(2), 187-217.
- Ladefoged, P. (2001). *Vowels and consonants: An introduction to the sounds of languages*. Oxford: Blackwell Publishers.
- Lee, C-H. (2010). Improving inter-rater reliability in oral proficiency test at college level. *Modern Studies in English Language & Literature*, 54(1), 367-387.
- Lowenberg, P. H. (2000). Assessing English proficiency in the global context: The significance of non-native norms. In H. W. Kam & C. Ward (Eds.). *Language in the global context: Implications for the language classroom* (pp. 207-228). Singapore: SEAMEO Regional Language Center.
- Lowenberg, P. H. (2002). Assessing English proficiency in the expanding circle. *World Englishes*, 21, 431-435.
- Luoma, S. (2004). *Assessing speaking*. Cambridge: Cambridge University Press.
- Mennen, I. (2006). Phonetic and phonological influences in non-native intonation: An overview for language teachers. *Queen Margaret University College speech science center working paper*, 9, 1-18.
- Munro, M., & Derwing, T. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45(1), 73-97.
- Munro, M., & Derwing, T. (2006). The mutual intelligibility of L2 speech. *Studies in Second Language Acquisition*, 28(1), 111-131.
- Ortega, L. (1999). Planning and focus on form in L2 oral performance. *Studies in Second Language Acquisition*, 21(1), 109-148.
- Purcell, E., & Suter, R. (1980) Predictors of pronunciation accuracy: A reexamination. *Language Learning*, 30(2), 271-287.
- Riazantseva, A. (2001). Second language proficiency and pausing. *Studies in Second Language Acquisition*, 23(4), 497-526.
- Rossiter, M. J. (2009). Perceptions of L2 fluency by native and non-native speakers of English. *Canadian Modern Language Review*, 65(3), 395-412.
- Schachter, S., Christenfeld, N., Ravina, B., & Bilous, F. (1991). Speech disfluency and the structure of knowledge. *Journal of Personality and Social Psychology*, 60(3), 362-367.

- Skehan, J. (2003). Task-based instruction. *Language Learning*, 36(1), 1-14.
- Sluijter, A. M. C., & Heuven, V. J. V. (1996). Acoustic correlates of linguistic stress and accent in Dutch and American English. *International Conference on Spoken Language Process*, 2, 630-633.
- Taylor, L. B. (2006). The changing landscape of English: Implications for language assessment. *ELT Journal*, 60, 51-60.
- Trofimovich, P., & Baker, W. (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.
- Ueyama, M. (2000). *Prosodic transfer: An acoustic studying of L2 English and L2 Japanese*. Unpublished doctoral dissertation, UCLA, Los Angeles.
- Wennerstrom, A. (1994). Intonational meaning in English discourse: A study of non-native speakers. *Applied Linguistics*, 15(4), 399-420.
- Wennerstrom, A. (2002). *The music of everyday speech*. New York: Oxford University Press.
- Willems, N. J. (1982). *English intonation from a Dutch point of view*. Dordrecht: Foris Publication

Seokhan Kang
#307, 10-1 Dong
Institute of Foreign Language Education
Seoul National University
599, Gwanak-ro, Gwankak-gu
Seoul, 151-748, Korea
82-2-880-7616
kangs45@snu.ac.kr
<https://sites.google.com/site/reall2research>

Received on October 15, 2012

Revised version received on December 4, 2012

Accepted on December 4, 2012