Linguistic Factors Affecting Evaluation of L2 Korean Speech Proficiency

Seung Hee Yang[1], *Minhwa Chung*[1, 2]

[1]Interdisciplinary Program in Cognitive Sciences, Seoul National University, Republic of Korea
[2]Department of Linguistics, Seoul National University, Republic of Korea
sy2358@snu.ac.kr, mchung@snu.ac.kr

Abstract

Much research attention has been directed to identify how native speakers perceive non-native speakers' oral proficiency. To investigate the generalizability of previous findings, this study examined segmental, phonological, accentual, and temporal correlates of native speakers' evaluation of L2 Korean proficiency produced by learners with various levels and nationalities. Our experiment results show that proficiency ratings by native speakers significantly correlate not only with rate of speech, but also with the segmental accuracies. The influence of segmental errors has the highest correlation with the proficiency of L2 Korean speech. We further verified the validity of this finding across different L1 backgrounds. Although phonological accuracy was expected to be highly correlated with the proficiency score, it was the least influential measure. Another new finding in this study is that the role of pitch and accent has been underemphasized so far in the non-native Korean speech perception studies. This work will serve as the groundwork for the development of automatic assessment module in Korean CAPT system.

Index Terms: CAPT (Computer-Assisted Pronunciation Teaching), Second Language Learning, Oral Proficiency Assessment, Non-native Korean Speech

1. Introduction

In second language (L2) acquisition, a growing number of researchers have emphasized the importance of assessing L2 speech proficiency based on judgments of comprehensibility, accentedness, and intelligibility [1][2][3][4][5][6][7]. They have studied what kinds of linguistic properties, such as phonetic accuracy, fluency, and grammar errors, are relatively crucial for native speakers' assessment under various task conditions.

This is a significant topic in language learning, since the results from these studies direct how L2 learners can achieve successful oral communication in the language. Indeed, the factors that may interfere with communication, and the degree to which they determine perceptual significance need to be identified for second language instructors, curriculum designers, and language learning software developers, since these standards will be manifested in the language teaching methodology. For instance, accentedness may not always spill over to lack of intelligibility. In a similar way, mispronunciation at the phonetic level may not necessarily interfere with communication. Some types of error influence the overall proficiency score more than others, and it is pedagogically important to obtain a better understanding of the correlation among the factors affecting listeners' perceptions.

This study has been conducted within the framework of CAPT (Computer-Assisted Pronunciation Teaching) system development for Korean. A CAPT system automatically assesses learners' speech and provides corrective feedback. In order to implement an automatic assessment module, it is necessary to examine how the task is carried out by human raters, and what criteria of judgment are used. For example, whether or not phonological and phonetic errors should be scored with equal weights in Korean, the context in which they cause miscommunication, and the degree of listeners' perceptual sensitivity need to be investigated.

It is noteworthy that previous literature on speech assessment considered L2-dependent factors in dealing with judgments of L2 English, German, Spanish, Japanese, Dutch and Chinese [1][2][3][4][5][6][7]. For example, in case of L2 English, due to the empirical evidence in favor of prosody as a factor of intelligibility, it was included as a variable in the experiment [2]. Studies on L2 Chinese investigated tonal realization patterns [6], whereas pitch accent was studied for L2 Japanese [4]. It remains open to question whether and to what degree such findings can be generalized to other linguistic contexts. In this study, L2-dependent factors for Korean were included as variables in order to identify the linguistic properties of Korean speech that are perceived as crucial for native evaluators.

Section 2 summarizes previous findings on non-native speech assessment and Section 3 proposes an improved experiment design. Section 4 describes the results of our experiment, followed by discussion and conclusion in Sections 5 and 6.

2. Related Studies

This section summarizes related studies on non-native speech assessment. The studies on L2 Korean will be examined in more detail in Section 2.2., before proposing an improved experiment for the current research in Section 3.

2.1. Non-native Speech Assessment

L2 education researchers have emphasized the importance of analyzing L2 speech with comprehensibility, intelligibility, and accentedness. According to Derwing and Munro's seminal work on accentedness and native speakers' comprehensibility [1][2], utterances that are perceived as heavily accented can be highly comprehensible. The finding showed that the degree to which learners approximate the native speaker norm does not necessarily measure how easily L2 utterances are understood. More empirical studies have examined phonological, temporal, lexical or grammatical correlates of L2 German [3], Japanese [4], Dutch [5], Chinese [6] and Spanish [7] comprehensibility.

2.2. Non-native Speech Assessment for L2 Korean

Table 1 shows the evaluation criteria employed by previous experiments that deal with L2 Korean. They assessed whether or not meaningful correlation can be observed between a fixed number of factors and proficiency scores [8][9][10][11][12]. The proficiency criterion refers to a holistic measure according to the rater's impression of proficiency in the utterance. Holistic measure of proficiency is distinguished from analytic measures since the raters rely on comprehensive impression across the entire utterance rather than paying attention to particular linguistic properties, such as fluency, phonology, or phonetics [13].

The filled circles in Table 1 indicate the factors that are highly correlated with the overall proficiency score according to each experimental results. There is no filled circle in the columns corresponding to [14] and [15] because these studies did not measure correlations between variables, but were interested in the change within a variable across time. Results in [14] concluded that fluency score improves for 6 months and starts degrading, while [15] found that all learners show different improvement patterns over time.

Table 1: *Evaluation criteria used in previous studies assessing non-native Korean speech* (○= *used as a variable*, ●=*used as a variable and found to be an important feature).*

Evaluation Criteria	[8]	[9]	[10]	[11]	[12]	[14]	[15]
Pitch	●						
Juncture	●		●				
Fluency	●			●	●	○	○
Segmental accuracy	○	●		○		○	○
Phonological accuracy	○						
Complexity				○		○	○
Hesitation					○		
Compre-hensibility					○		○

Table 2: *Comparison of experimental set-ups in previous studies assessing non-native Korean speech (NS = Native Speaker).*

Study	Task	No. of Speakers (L1)	No. of Raters
[8]	Read 10 sentences	33 (Chinese)	3 (NS)
[9]	Read 210 sentences	24 (Japanese)	4 (NS)
[10]	Read 25 sentences	24 (Chinese)	10 (NS)
[11]	Interview after watching Mr. Bean movie	130 (Chinese, French, Japanese, English)	6 (NS)
[12]	Open question answering, picture story telling	37 (Thai)	5 (NS)
[14]	Interview after watching a movie	20 (Chinese)	3 (NS)
[15]	Talk about a given topic	3 (English, Chinese, Russian)	1 (NS)

All studies that measured fluency as the evaluation criteria agree that it is a useful measure which highly correlates with native listeners' perception of speech proficiency [8][9][10]. The correlation was shown to be stronger than segmental accuracy [8][11]. However, segmental accuracy, including all substitutions, deletions, and insertions, is still important according to [9], while number of juncture insertion is an important consideration in [10].

The causes of these inconsistencies can be partly explained by the nature, design, and purpose of different experimental set-ups, which is summarized in Table 2. For example, L1 specificities may have introduced some disagreements, and it may be the case that segmental accuracy is more crucial for speakers with Japanese background, while suprasegmental features are more important for speakers with Chinese background. Moreover, whether the evaluation sample was read speech or spontaneous speech would play a big role in the assessment, as it introduces other factors such as orthographical influence and the learners' knowledge of the words. For evaluating read speech prompts, they used proficiency score as a criterion [8][9][10], while the studies using spontaneous speech included comprehensibility or complexity measures [11][12][14][15]. The disagreements raise further research questions and call for an improved experiment that can clarify which areas of language (fluency, juncture, segmental, phonological accuracy) influence native speakers' judgments of L2 Korean.

Moreover, some variables have been left out or have been insufficiently considered, which can be a limitation of the experiments. For instance, most of them did not measure the effect of pitch and stress errors, which have been influential factors in L2 evaluation [1][2][4]. Also, the coverage of phonological rules is not comprehensive. One of the characteristics of Korean is its usage of phonological rules, i.e., pronunciation at the surface level is changed by certain conditions in the underlying representation. Several studies have reported that learners of Korean are pronouncing the segments according to their underlying representation, and phonological rules are not realized [16][17][18][19]. Therefore, the extent to which phonological accuracy affects the assessment of L2 Korean speech needs to be thoroughly investigated. However, only a portion of the phonological phenomena has been covered in [8], and it is necessary to design an experiment that is comprehensive in scope. The next Section will discuss in more detail what the missing phenomena were, and how we propose to improve the experiment.

3. Experiment Design

This section describes our methodology and what improvements have been made compared to the previous studies. It elaborates on how we composed the reading prompts, variables, evaluation method, speakers, and evaluators.

3.1. Reading prompts

50 speakers were given 100 sentences to read. The sentences used everyday vocabulary from L2 Korean text books, such as "How many times have you been to Korea?" and "I usually eat dinner when I go home." Since spontaneous speech is closer to the real communication phenomenon, using read speech corpus can have limitations. However, canonical pronunciation is predefined in read speech, which can be an

advantage for discovering error patterns, and also for conducting a research with the beginner level speakers, whose canonical form of the utterance are often impossible to identify.

Moreover, using read speech prompt enables a comprehensive analysis of phonological accuracy. For this study, we used 50 sentences containing carefully selected phonological rules. As shown in Table 3, there are 264 instances of phonological rules occurring in the prompt. Five common phonological rules for both cross and within-morphemes are included. For example, tensification rule in the word "worry" [gʌk͡z͈ʌŋ] occurs within morpheme, whereas the aspiration rule in the word "would like" [d̥͡ʒok^{h}es͈u͡mni̥a] occurs across morpheme. The frequencies are balanced to cover diverse linguistic phenomena. Regarding sentence types, the 50 sentences consist of 42 statements and 8 questions.

Table 3: *Distribution of phonological rules occurring in the 50 sentences used in this experiment.*

Phonological Rule Type	Cross-Morpheme (freq.)	Within-Morpheme (freq.)
Lenition	82	5
Tensification	33	13
Nasalization	10	25
Aspiration	26	4
Palatalization	2	3
Total	158	106

3.2. Selection of variables

The purpose of the current investigation is to examine the generalizability of previous findings in [8][9][10][11][12] and resolve their disagreements. We also examine whether and to what degree the correct realization of phonological rules affect L2 Korean perception. In order to answer these research questions, the following five variables have been defined: segmental accuracy, phonological accuracy, pitch and accent, fluency, and holistic impression of proficiency. Upon listening to each sample, the raters used 1-5 Likert scale (5: perfect, 4: good, 3: acceptable, 2: poor, and 1: very poor) to evaluate each variable.

Figure 1 illustrates the difference between phonological and segmental accuracies. The top row is the pronunciation according to the underlying form of the characters, before the phonological rules are applied, whereas the bottom row is the canonical pronunciation after correct application of the rules. In this example, the phonological phenomenon include aspiration, tensification, and nasalization rules, in which case the correct→realized substitutions of k^{h}→g̥, s͈→s, m̄→p̄ at these positions will be counted as phonological errors. All the positions where the rules occur and their types are marked in advance, so that the raters know which errors to listen to. All other substitutions are considered as segmental errors.

1) Proficiency: As employed in previous studies [8][9][10], proficiency was rated by evaluators' impressionistic and holistic judgments of the overall utterance in the scale of 1 to 5, without paying attention to specific linguistic features. For example, even if a part of an utterance digresses from the canonical, they can assign high scores if it is perceived as acceptable.

2) Fluency: The evaluators rated fluency based on rate of speech, juncture, pause, and filled pause. For example, novice learners tend to speak slowly and pronounce each syllable separately, which is not observed in native speech. This would discount the fluency score.

3) Phonological accuracy: All syllables where phonological rules occur and their types are marked in advance for the five different phonological phenomenon, which are listed in Table 3. In this way, the raters know which errors to listen to. They counted the number of errors and gave scores based on the count.

4) Segmental accuracy: The raters phonetically transcribed all segments and rated segmental accuracy according to the number of mismatch between the canonical and realized pronunciations.

5) Pitch and accent: The evaluators judged the appropriateness of pitch and accent realized at lexical and sentential levels. For example, if a question is perceived as a statement due to inappropriate pitch realization, the utterance will receive a low score.

Note that the raters had a prior knowledge of the read prompts and therefore, comprehensibility was not included as a variable. That is, the degree of effort required by raters to understand an utterance could not be independently measured in this study by the nature of read speech task. This is also consistent with the previous studies that evaluated read speech by holistic impression of proficiency, instead of comprehensibility.

Moreover, only one pronunciation per word is defined as the canonical form in this experiment. There are certainly variants and regional variations that are also recognized as acceptable pronunciations, which means that some may not be counted as an error. However, we did not allow multiple correct answers in this experiment because there is no consensus on what counts as an 'acceptable variance,' and would cause confusion in the scoring process. In case when multiple pronunciation rules can be applied to the same phonemic context, we take the most common pronunciation as the canonical pronunciation. In addition, predefined standard Korean pronunciation exists according to the National Institute of Korean Language, and is the form of Korean that is accepted as a national norm. Therefore, considering that the purpose of current research is a pedagogical application, it seems desirable to keep the correct reference as the gold standard. Such variations that are observed in native speech will be considered with lower priority in later feedback stage.

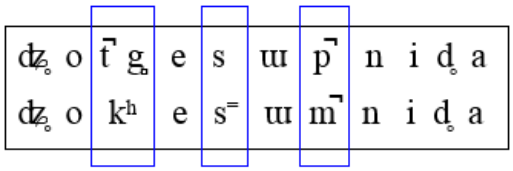

Figure 1: *Pronunciations before and after the application of phonological rules occurring in the phrase [d̥͡ʒok^{h}es͈u͡mni̥da] (would like). The highlighted boxes show where phonological errors may occur, while mispronunciations of segments outside the boxes are counted as segmental errors.*

3.3. Speakers

50 speakers of L1 Mandarin Chinese, Japanese, Cambodian, Vietnamese, and Filipino residing in Korea are included. Gender, age, and learner levels are balanced from beginner to advanced. The speaker age ranges from 18 to 60, with the average of 24. Their average length of residence in Korea is 1.5 years, ranging from 2 months to 6 years.

3.4. Evaluators

Each utterance was scored by four native Korean graduate students in Seoul National University with knowledge in Korean phonetics and phonology. Since phonological accuracy was included in the evaluation criteria, it was necessary to recruit raters with detailed knowledge in Korean phonological rules in this evaluation task.

The evaluators practiced scoring with the established guidelines to ensure inter-rater consistency. Before the four raters could officially start scoring, we made sure that the correlation was consistent on the first 50 utterances for training purposes. To ensure consistent scoring quality, we held biweekly training and discussion sessions for monitoring inter-rater consistency throughout the scoring and annotation period, which took about five months.

3.5. Summary of methodological improvements

Compared to the previous experiments, improvements have been made in the coverage and depth of variables and speaker diversities. First, we have increased the number and the scope of variables. Most previous researches mentioned in Section 2.2 studied correlations between two or three variables [9][10][11][12], whereas our study looks at them more comprehensively. We expect this will enable a direct comparison among the variables, and thereby resolve the existing disagreements. Second, the coverage of phonological rules has been increased by adding aspiration and palatalization error types, in addition to lenition, nasalization, and tensification. We believe this is important because phonological rule is a pedagogically meaningful area where the learners may need explicit instruction. Third, we attempt to reduce the disagreements arising from L1 effect by including participants of various backgrounds, with L1 Mandarin Chinese, Japanese, Cambodian, Vietnamese, and Filipino, which is more diverse than other previous experiments with L2 Korean.

4. Experimental Results

The four raters demonstrated general agreement ($\alpha = 0.88$) on the proficiency rating task over 2,500 utterances, suggesting that they share similar intuitive notion of what it meant by holistic impression of proficiency in L2 Korean speech. Inter-rater reliability was calculated by using Cronbach's alpha. The coefficients reported in the previous studies confirm that the results are reliable ($\alpha = 0.82$ [4], 0.88 [9], 0.89 [10], 0.74 [12]). The mean and standard deviation of proficiency scores are 2.94 and 0.98, respectively, and their distribution is summarized in Table 4. In the following analyses, all raters' scores were averaged to derive a single score for the perceived proficiency of each utterance.

Table 4: *Proficiency score distribution for 2,500 utterances, each rated by four native speakers.*

Score	1	2	3	4	5	Total
No. of Utterances	1,273	2,906	3,388	1,971	462	10,000

4.1. Linguistic correlates of speech proficiency

We conducted a set of correlation analyses to examine how proficiency rating scores were related to the four linguistic variables defined in the previous section. The results of the linguistic influences on their proficiency evaluation are summarized in Figure 2. All variables are strongly correlated with proficiency scores. Among the variables, proficiency was most strongly correlated with segmental accuracy ($r = 0.81$) and fluency ($r = 0.80$), and relatively weakly correlated with pitch and accent ($r = 0.76$) and phonological accuracy ($r = 0.74$). All correlations are statistically significant ($p<0.0001$). In other words, speech with higher proficiency ratings was comprised of fewer segmental errors, and was fluently spoken with an appropriate rate of speech. This suggests that the raters similarly relied on segmental and fluency information during their proficiency judgments.

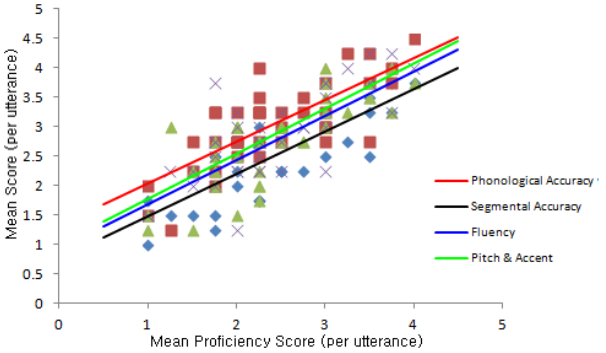

Figure 2: *Linguistic correlation with proficiency scores according to Pearson measure. Correlation is the highest in the order of segmental accuracy ($r = 0.81$), fluency ($r = 0.80$), pitch and accent ($r = 0.76$), and phonological accuracy ($r = 0.74$).*

Similar to the correlation analyses in previous studies, we confirm that speech rate is a significant predictor of speech proficiency. However, it is a new finding in this study that segmental accuracy is an even better predictor of proficiency than speech rate. Although the difference between them is small, the finding is significant because most previous studies with L2 Korean did not take segmental accuracy to be an important measure.

Another important finding in this experiment is that pitch and accent is correlated with proficiency ($r = 0.76$). Most studies with L2 Korean did not include pitch and accent as a variable, which can be partially explained by the fact that Korean is a syllable-timed language and it is easy to assume that the difference in pitch and accent would not be perceptually significant. However, the experimental result in the current study is contrary to the expectation. In fact, pitch and accent is shown to have even higher correlation than phonological accuracy, which was initially predicted to be a challenging area according to the related works in L2 Korean. Therefore, it is surprising that pitch and accent plays a significant role in L2 Korean perception, to an equivalent or higher degree than phonological accuracy.

A possible explanation for this new observation lies in L1 interference by Chinese speakers. The influence of tonal pitch can result in pitch lowering at the end of accentual phrases, which may be perceived as accented and unnatural for the native listeners [8]. In fact, the average score for pitch and accent is lower for L1 Chinese, 3.04, than L1 Japanese, 3.18 with the standard deviations 0.99 and 1.07, each respectively. An independent group t-test was performed to compare the averages between the two groups. The t-statistic was significant at the 0.05 level ($p=0.0071$). This stresses the importance of considering L1 dependent factors in experiment design. It seems that such L1-dependent factors have been underemphasized so far, and pitch and accent have more sensitivity in Korean speech perception than was predicted.

4.2. L1 dependent segmental variation patterns

The experiment results show that segmental accuracy affects the proficiency score more strongly than other factors. We further verified this finding across L1 backgrounds. Figures 3, 4, 5, and 6 show that there are diversities in segmental production patterns across L1 backgrounds. Southeast Asian languages here include Filipino, Cambodian, and Vietnamese, which belong to Austronesian language group. The patterns are organized into five groups (coda deletion and insertion, detensification, aspiration, and vowel epenthesis), which are the salient mispronunciation patterns in non-native Korean speech [20][21]. Chinese and Japanese L1 speakers show higher coda deletion errors than Southeast Asian learners (Figure 3). Detensification and coda insertion error is characteristic of Chinese learners (Figure 4), while vowel epenthesis is characteristic of Japanese speakers (Figure 6). For Southeast Asian speakers, aspiration error is more frequent than other groups (Figure 5). The analysis demonstrates that all L1 groups have salient segmental error patterns that contribute to lower proficiency scores, and confirms the new finding of this research across learners' L1 backgrounds.

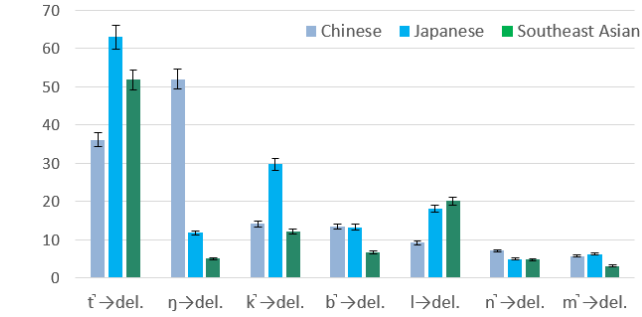

Figure 3: *Variations in coda deletion error patterns across learners' L1 backgrounds (canonical pronunciation→realized pronunciation, del. = "deletion")*

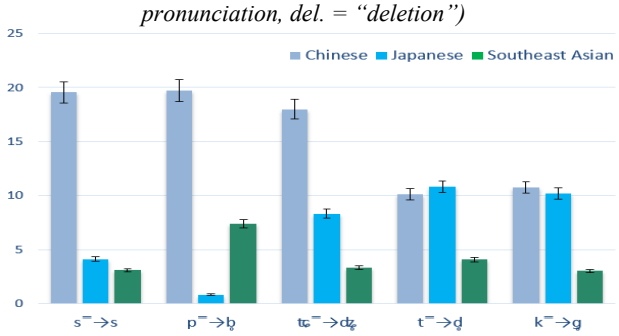

Figure 4: *Variations in detensification error patterns across learners' L1 backgrounds (canonical pronunciation→realized pronunciation)*

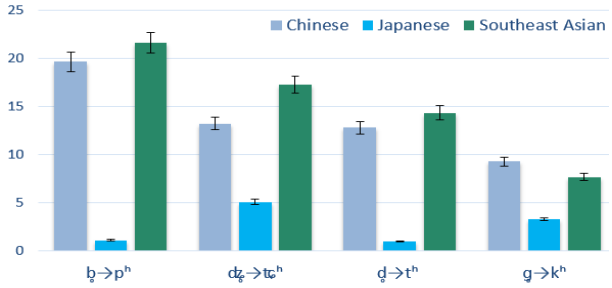

Figure 5: *Variations in aspiration error patterns across learners' L1 backgrounds (canonical pronunciation→realized pronunciation)*

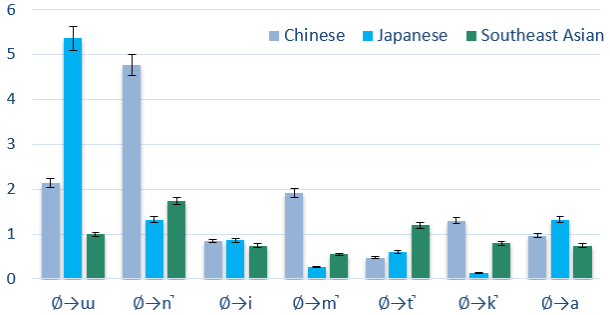

Figure 6: *Variations in insertion error patterns across learners' L1 backgrounds (canonical pronunciation→realized pronunciation, ∅ = "no segment")*

5. Discussion

In Section 2, the research questions were formulated as follows:

1. Which linguistic features predict the holistic speech proficiency of L2 Korean speech?

2. To what extent does the correct realization of phonological rules affect L2 Korean speech proficiency?

In response to the first research question, the results indicate that the effects of segmental accuracy, fluency, pitch and accent, and phonological accuracy are all positively correlated with L2 speakers' oral proficiency scores obtained from native listener judgements, in the order of their importance. Therefore, segmental accuracy is significantly predictive of L2 Korean proficiency judgment. This resolves the disagreement regarding the relative importance of segmental and suprasegmental features in the previous researches. That is, there was a disagreement whether fluency, segmental accuracy, or both are strong predictors of L2 Korean speech proficiency. We showed that both are highly correlated with proficiency, and segmental accuracy is in fact, more strongly correlated.

As for the second research question, we found that phonological accuracy is least correlated with the proficiency scores. This is interesting because it is contrary to the previous studies' prediction that phonological accuracy is an L2-specific property in Korean that learners would find difficult. A possible explanation is that not all mispronunciations at the phonological level cause difference in meaning and therefore, does not influence raters' impression of proficiency. For example, pronouncing the word [tɕʰukʰa] "to congratulate" without applying the phonological rule to aspirate and reading it as the underlying text form, [tɕʰuk̚ha], does not change the word into a different meaning. The ways in which phonological accuracy affect raters' perceptual impression can be a subject of future research.

There is a room for improvement in the current study regarding the independence of individual variables, especially between the segmental and pitch accuracies. For example, syllables with tense consonants in Korean are realized with higher pitch than the lax counterparts [22], leading to interaction effect between segmental and pitch variables. In such cases, it can be difficult to identify if the error originates from phonetic or prosodic category. In the future studies, it seems desirable to make more efforts to minimize such effects.

In the future works, we plan to extend this study to spontaneous speech. The findings of read speech may not necessarily generalize into spontaneous speech. We believe that using read speech recording has an advantage in that the reference pronunciation is known, and therefore, error patterns can be efficiently discovered. However, spontaneous speech is closer to natural communication settings, and we plan to extend this study to spontaneous speech in the future works.

6. Conclusion

Using the speech produced by fifty L2 learners of Korean, the current study investigated the linguistic correlates of oral proficiency. Certain errors are believed to entail more perceptual value than others. According to the results in the correlational analysis, our findings for L2 Korean were generally consistent with the pervious literature, in that fluency score is a good measure of oral proficiency, including speech rate, juncture, and other temporal features. Furthermore, positive correlation was found between accuracies in Korean phonological rules and oral proficiency scores, which also echoed previous literature that highlighted the importance of assessing L2-specific variables.

The new finding in this study is that segmental accuracy demonstrates the highest correlation with the proficiency scores. Moreover, native listeners are more sensitive to pitch and accent than it was predicted, and deserve more attention in L2 Korean studies. In contrast, phonological accuracy is relatively less important. Building on the previous studies concerning L2 acquisition, the results presented here postulate a verified hypothesis on the oral assessment for L2 Korean. In future works, these findings will not only be extended to spontaneous speech, but also will serve as the ground for the development of automatic assessment module in Korean CAPT system.

7. Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, the Republic of Korea (No. 2015R1D1A2A01061378).

8. References

[1] J. Munro and T. Derwing, “Foreign accent, comprehensibility, and intelligibility in the speech of second language learners,” *Language learning,* vol. 45.1, pp. 73-97, 1995.

[2] T. Derwing, M. Tracey and J. Munro. "Accent, intelligibility, and comprehensibility," *Studies in second language acquisition*, vol. 19.01, pp. 1-16, 1997.

[3] O'Brien, “L2 Learners’ Assessments of Accentedness, Fluency, and Comprehensibility of Native and Nonnative German Speech,” *Language Learning*. vol. 64.4, pp. 715-748, 2014.

[4] Saito, Kazuya, and Akiyama, “Linguistic correlates of comprehensibility in second language Japanese speech,” *Journal of Second Language Pronunciation*. 2017.

[5] A. Neri, C. Cucchiarini and H. Strik, “Segmental errors in Dutch as a second language: how to establish priorities for CAPT,” *iCALL Symposium,* Venice, Italy, June 2004.

[6] N. Chen, D. Wee, T. Rong, B. Ma and H. Li, “Large-scale characterization of non-native Mandarin Chinese spoken by speakers of European origin: Analysis on iCALL,” *Speech Communication*. vol. 84, pp. 46-56, 2016.

[7] K. McBride. “Which features of Spanish learners' pronunciation most impact listener evaluations?” *Hispania*, vol. 98, pp. 14-30, 2015.

[8] M. Jung, “Strategic Pronunciation Education for Chinese Learners of Korean, Korean Language Studies,” *The Society of Korean Language and Literature,* vol. 38. pp. 345-369, 2008.

[9] H. Hong. “The relationship between segmental production by Japanese learners of Korean and pronunciation evaluation,” *Phonetics and Speech Sciences*, vol. 6(4), pp. 101-108, 2014.

[10] S. Hong. “Juncture Patterns in Chinese learners of Korean,” M.A. Thesis, Seoul: Korea University. 2016.

[11] Y. Kim, J. Nam, S. Lee, S. Lee. “A Study on the Correlation Between Proficiency and Productive Ability of Learners of Korean,” *The Society of Korean Language and Literature,* vol. 164, pp. 209-244. 2013.

[12] S. Sayamon, “The Influence of Learning Contexts on Speaking Development of Intermediate Thai Learners of Korean - Focusing on Fluency and Pronunciation,” Ph.D. dissertation, Seoul: KyungHee University, 2016.

[13] C. J. Alderson, D. Wall, and C. Claphaim. *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press, 1996.

[14] A. Kim, “A Study on Korean Language Learner's Speaking Ability Development - Focusing on Chinese Learners Residing in Korea,” Ph.D. dissertation, Seoul: KyungHee University, 2017.

[15] B. Lee. “A study on the dynamic development of complexity, accuracy and fluency in Korean learners' writing and speaking,” Ph.D. dissertation, Seoul: Yonsei University, 2016.

[16] C. Yoo. “Comparison of Phonology Between Korean and Chinese for Korean Pronunciation Education,” M.A. Thesis, Changwon: Changwon University, 2012.

[17] K. Chang. “(A) comparative study on phonology the phenomena between Korean and Chinese,” M.A. Thesis, Wonju: Gangneung Wonju National University, 2014.

[18] T. Chung. “(A) comparative study of the Korean and Chinese phonological system: focusing on the exploration of the pronunciation teaching program,” M.A. Thesis, Nonsan: Konyang University, 2014.

[19] S. Lee, J. Kim, and J. Chang., “Design and Construction of Speech Corpus for Korean as a Foreign Language (L2KSC),” *The Journal of Chinese Language and Literature*, vol. 33, pp. 35-53, 2005.

[20] S. Yang, H. Ryu, and M. Chung. “A Corpus-based Analysis of Korean Segments Produced by Chinese Learners,” *Proceedings of APSIPA*, Hong Kong, pp. 583-586, Dec. 2015.

[21] H. Hong, S. Kim, and M. Chung. “A corpus-based analysis of Korean segments produced by Japanese learners,” *Proceedings of SLaTE*, Porto, Portugal, pp. 189-192, June 2013.

[22] J. In and C. Seong, “Characteristics of AP Tonal Patterns & Slopes Produced by Chinese Learners of Korean,” *Phonetics and Speech Sciences*, vol. 5(3), pp. 47-54, 2013.