Native and nonnative evaluators' assessments of two proficiency levels of English suprasegmentals[*]

Eunkyung Sung
(Cyber Hankuk University of Foreign Studies)

Sung, Eunkyung. 2011. Native and nonnative evaluators' assessments of two proficiency levels of English suprasegmentals. *Studies in Phonetics, Phonology and Morphology* 17.3. 405-425. This study investigates native and advanced-level nonnative evaluators' assessments of second language (L2) speech from Korean speakers at two English proficiency levels, focusing on suprasegmental features such as intonation, rhythm and word stress, as well as on overall proficiency. The results show that the ratings done by native evaluators are significantly lower than those done by nonnative evaluators. However, discrepancies in the assessments of the two groups of evaluators are not consistent across the two groups of assessed speakers. That is, the differences between the assessments of the two groups of evaluators are statistically significant only for the speech samples from the higher proficiency group. Although advanced-level nonnative evaluators were able to recognize completely inappropriate pitch patterns in L2 speech, slightly inappropriate pitch patterns were not salient to them. Furthermore, for the higher proficiency group, the significant differences between the assessments of the two groups of evaluators appeared in intonation and overall proficiency. Considering that suprasegmental features are more salient than individual segments in the distinction of foreign accents, and that instruction in suprasegmental factors could result in significant improvements in L2 speech proficiency, more attention should be paid to suprasegmental features in pedagogical settings even for L2 speakers at high proficiency levels. At the conclusion of the study, suggestions are presented for future research that could address the comparison of L2 suprasegmental assessments of evaluators from various language backgrounds, as well as the relationship between instruction-driven development of L2 perception and production. **(Cyber Hankuk University of Foreign Studies)**

Keywords: English, assessments, native evaluators, nonnative evaluators, suprasegmental features, two proficiency levels

1. Introduction

It has been shown that poor suprasegmental skills can have a more devastating effect on communication and can make conversation more frustrating than poor pronunciation of individual sounds (Anderson-Hsieh et al. 1992, Derwing and Rossiter 2002, Hahn 2004, Boula de Mareuil and Veru-Dimulescu 2006). Inappropriate timing and stress patterns often lead

to intelligibility[1] deficit, and intonation errors may contribute to the perception of foreign accents (Pickering 1999, Jilka 2000, Hahn 2004). Although a foreign accent has both segmental and suprasegmental characteristics, the majority of second language (L2) research has focused on segments, while paying little attention to the contribution of prosodic dimensions (Boula de Mareuil and Veru-Dimulescu 2006). Recently, however, language teachers have shifted the focus of their pronunciation teaching towards the inclusion of suprasegmental features along with segmental sounds in order to improve general L2 proficiency of students (Celce-Murcia et al. 2010).

Although the role of suprasegmental features in L2 speech has been more widely recognized, there has not been much research to examine the relations of assessments between native and nonnative evaluators focusing on suprasegmental aspects of L2 speech. Furthermore, very few studies have been carried out to investigate the assessments between native and nonnative evaluators involving the L2 speech from two groups of speakers at different proficiency levels.

The aim of this study is to find whether native and nonnative evaluators differ in their evaluation of L2 speech with regard to suprasegmental features such as intonation, rhythm, and word stress, as well as overall proficiency. Furthermore, this study investigates the assessments of English utterances produced by two groups of Korean speakers at different English proficiency levels, a higher proficiency and a lower proficiency level. Twenty Korean speakers at the two different L2 proficiency levels read an English passage, and their readings were recorded and assessed by two groups of evaluators – eight native English speaking evaluators and eight advanced-level nonnative English speaking evaluators.

The following research questions are to be examined in this paper: Do native English speaking evaluators and evaluators who are advanced-level L2 learners of English assess suprasegmental features of L2 speech differently? If so, are the discrepancies in assessments between native and nonnative evaluators more clearly shown in speech from L2 speakers at a lower proficiency level or from those at a higher proficiency level? Finally, for which features of L2 pronunciation - intonation, rhythm, word stress, and overall proficiency – are the assessments more varied?

Previous research has shown that ratings of pronunciation by native speakers can usually be regarded as reliable (Flege 1984, Cucchiarini et al. 2000). Therefore, in this study L2 learners' assessments are evaluated on the basis of their agreement with the assessments by native speakers.

---

[1] The terms, intelligibility and comprehensibility, need to be distinguished. The former commonly refers to the extent to which an utterance is actually understood by a listener, while comprehensibility typically refers to a listener's perception of the amount of effort involved in understanding a particular nonnative speaker (Munro and Derwing 1995, Derwing and Munro 1997).

2. Literature review

2.1 Suprasegmental features in L2 speech

Nonnative accentedness may be detected through suprasegmental aspects as well as individual segments. Previous research has shown that in the perception of foreign accents, suprasegmental features are more salient than, or at least as salient as, segments (Anderson-Hsieh et al. 1992, Derwing and Munro 1997, Tajima et al. 1997, Derwing and Rossiter 2002, Hahn 2004, Boula de Mareuil and Veru-Dimulescu 2006). Anderson-Hsieh et al. (1992) investigated ESL teachers' impressionistic judgments of nonnative speakers' production in terms of segments, syllable structure and prosody. They found that prosody played the most important role in the perception of foreign accents. Derwing and Munro (1997) concluded that the improvement in nonnative speakers' comprehensibility by native evaluators was more affected by grammatical and prosodic proficiency than by phonemic proficiency. Tajima et al. (1997) aligned Mandarin English with native English timing patterns by using LPC resynthesis, and found a significant increase in intelligibility. They argued that nonnative speakers would greatly benefit from training program focusing on various prosodic aspects of their speech. Hahn (2004) examined native English speakers' reactions to nonnative speakers' primary stress placement in English, and found that incorrect placement or complete absence of stress led to poor assessments of the nonnative speakers. Boula de Mareuil and Veru-Dimulescu (2006) explored the relative strength of the two aspects of pronunciation (i.e., phonemic and prosodic aspects) in the perception of Spanish-accented Italian and Italian-accented Spanish. They argued that the articulation of phonemes helped identify the speakers' first language (L1), but that prosody provided slightly more reliable indicator.

Previous studies have examined phonetic features such as speech rate, pause, word stress, and pitch pattern or intonation in assessing suprasegmental aspects of L2 speech. Relatively slow speech is often considered an indication of foreign accent (Derwing and Munro 2001). Anderson-Hsieh and Venkatagiri (1994) and Pickering (1999) showed a qualitative difference in both placement and length of pauses between native and nonnative speech. Trofimovich and Baker (2006) explored the relation between foreign accent and five prosodic elements (i.e., stress timing, peak alignment, speech rate, pause frequency, and pause duration). Acoustic analyses and listener judgments revealed that speech rate, pause frequency, and pause duration more strongly contributed to the presence of foreign accent than the other elements.

Nonnative speakers' word stress errors are likewise related to comprehensibility. Bond (1999) indicated that misplaced word stress was more perceptually salient to native evaluators than mispronounced phones. Field (2005) found that the comprehensibility of both native and nonnative

evaluators was deteriorated by the shift of lexical stress. Lee et al. (2006) studied the production of unstressed English vowels by L2 learners with respect to the four phonetic features, fundamental frequency (F0), intensity, duration, and vowel quality reduction. They found that Korean speakers were native-like in terms of F0, but that the intensity, duration, and vowel quality of their unstressed vowels were different from those of native English speakers' production.

Intonation also plays a key role in the comprehensibility of nonnative speech to native speakers as well as in the understanding of nonnative speakers' intent. Pickering (1999) reported that nonnative speakers were not consistently able to use the proper intonation patterns found in the native speakers' discourse. Using low-pass filtered stimuli with monotonous intonation, Jilka (2000) showed that intonation was the most important prosodic factor contributing to foreign accent among other prosodic features such as rhythm or speech rate.

A focus on suprasegmental features of L2 speech is important in the view of pedagogy. Elliott (1997) demonstrated that the teaching of segmental aspects did not lead to the significant improvement in communication ability. Derwing et al. (1998) used three instruction types (i.e., segmental accuracy, general speaking and prosodic factors, and no specific pronunciation instruction) for L2 learners. They found that speakers who had had instruction emphasizing prosodic features could improve comprehensibility in spontaneous speech. In the study of age, motivation, and instruction on L2 phonology, Moyer (1999) varied the type of phonological feedback, and found that subjects who received suprasegmental training performed closer to native level.

Much previous research has included suprasegmental factors such as rate, pause, lexical stress, and various aspects of pitch in order to analyze the deviation of L2 speech from the native norm. Furthermore, previous studies have relied heavily on native raters. However, there have been few attempts to compare the judgments of native and nonnative evaluators with respect to L2 suprasegmental aspects of speech. Since in Korea, nonnative speakers often teach and evaluate L2 speech in various instructional settings, it is worth examining the assessment differences between native and nonnative evaluators.

2.2 Assessments of L2 speech by native and nonnative evaluators

In assessing L2 speech, previous studies reported native and nonnative evaluators are very adept at noticing when speech is different from their own variety. Flege (1984) found that phonetically untrained native evaluators could identify nonnative speakers based on very short samples of speech. Native listeners were extremely sensitive to nonnative speech because they relied on various cues, segmental variation and suprasegmental factors. Flege (1988) posited that nonnative evaluators

showed similarities with native evaluators in detecting foreign accents of L2 speech based on the lack of statistical differences between native and nonnative assessments. Flege (1984) and Thompson (1991) independently proposed that linguistically experienced nonnative evaluators were more dependable than inexperienced ones in judging intelligibility of foreign language learners. van Wijngaarden (2001), Bent and Bradlow (2003), and Jongman and Wade (2007) observed that the nonnative speech was more intelligible to nonnative evaluators than to native evaluators.

Previous research on the relationship between native and nonnative assessments of L2 speech has revealed inconsistent results with respect to correlation between the two groups of evaluators. Rhee and Park (2003) compared assessments of an expert nonnative group, an inexperienced nonnative group, a native group, and an automatic speech recognition system. They found that the highest correlation appeared in the assessments of the two nonnative groups (r=.98), and that the correlation of the assessments of the native and nonnative groups was also high (r=.92). Furthermore, the correlation of assessments between human beings and machines was not low (r=.72). Yun (2009) investigated the characteristics of L2 pronunciation evaluation by L2 raters, and found that the source of inter-rater differences was neither random nor arbitrary. The L2 phones and phonotactic constraints which do not exist in the L1 sound system seemed to be difficult to evaluate correctly.

Lee (2010) revealed a high correlation between native and nonnative evaluators in a college level oral proficiency test in general. However, some discrepancies were shown in the aspects of grammar and pronunciation. Kang and Rhee (2011) studied the role of suprasegmental features (e.g. F0, speech rate, pause) in terms of the intelligibility of L2 speech to native Korean and English speaking listeners. They indicated that the assessment of intelligibility was different for the two groups and that the discrepancies arose from their perception of L2 suprasegmental features. Sung (2011) compared native and nonnative listeners' assessments of L2 speech, and found that nonnative listeners' ratings were significantly higher than those of native listeners for three suprasegmental features including intonation, word stress, and reduced speech. There were no significant differences in the assessments between the two listener groups for individual sounds.

The amount of research investigating the differences in L2 speech assessment between native and nonnative evaluators focusing on suprasegmental features is scarce. Furthermore, the results of previous studies which investigated the relationship between native and nonnative assessments have not been consistent. Moreover, little research has been done involving native and nonnative evaluators' L2 speech assessments of two groups of L2 speakers divided by their L2 proficiency levels.

3. Methods

3.1 Participants

3.1.1 Speakers

Twenty female Korean learners of English at two different proficiency levels (ten in the higher L2 proficiency group and ten in the lower L2 proficiency group) took part in the study by getting their readings of an English passage recorded. They were all graduate students at H University in Seoul. The participants' general English proficiency level was determined based on their self-reported scores of English proficiency tests. The ten lower proficiency level students majored in Chinese, Korean Education, or Music Education. Seven of them had taken the TOEIC exam (Test of English International Communication), and their average score was 767, ranging from 645 to 840. The other three participants had never taken any internationally authorized English tests. The ten higher proficiency level students were all English Education majors. All of these participants had taken the TOEIC exam and their average score was 931, ranging from 900 to 970. The 95% confidence intervals showed no overlap between the two groups in terms of mean TOEIC scores, suggesting distinct proficiency differences.

3.1.2 Evaluators

Two groups of evaluators participated in this study. One group consisted of eight female Korean learners of English who were graduate students at H University in Seoul, and the other group was comprised of eight native speakers of English who were instructors in the English department at the same university. All of the Korean learners of English were English Education majors, and their English proficiency was high, with TOEIC scores ranging from 880 to 990 (mean score 935). The Korean evaluators were enrolled in an English pronunciation course during the evaluation experiment, and had already learned the basics of phonetics as well as specific aspects of English pronunciation quality and various methods of teaching English pronunciation. They did not show any difficulty understanding lectures in English, or communicating with their peers and the instructor in English. Thus, all the Korean learners of English were rated as advanced-level learners of English. None of the Korean evaluators participated as subjects in the production phase of reading the English passage.

3.2 Materials

3.2.1 Diagnostic passage

Korean speakers' ability to use appropriate English stress and intonation was assessed based on the reading of a diagnostic passage. The diagnostic passage, a modified version of the one found in a course book (Celce-Murcia et al. 2010: 481), was composed of 18 sentences including 7 interrogative sentences (see Appendix A). The 7 interrogative sentences consisted of 4 yes/no questions and 3 wh-questions. Although a reading of the passage did not provide spontaneous pronunciation, it allowed the evaluators to obtain a sampling of various suprasegmental aspects of L2 speech errors.

3.2.2 Assessment form

In order to assess the L2 speech, an assessment form, which was a modified version of one found in the aforementioned course book (Celce-Murcia et al. 2010: 487), was used. The assessment form consisted of four pronunciation features: intonation (yes/no questions, wh-questions, statements), rhythm, word stress, and overall proficiency. Six items (including three intonation items) were measured on a five-point scale, with 5 representing native-like pronunciation and 1 indicative of a strong foreign accent. Below each of the five-point scale measurement for each item was provided a space for evaluators' comments (see Appendix B).

3.3 Procedure

3.3.1 Recording

Speech recording sessions were conducted one-on-one in a quiet office. The participants were asked to read the diagnostic passage at a natural speed, and their production was recorded to a computer using a GOM recorder and a headset (Plantronics DSP500). The participants were allowed time to practice before recording the passage to be sure that they knew every word in the passage. All speakers filled out a background questionnaire.

3.3.2 Assessment

Two groups of evaluators, respectively made up of eight native English speakers and eight native Korean speakers, listened to and rated the recorded speech independently. In order to evaluate the L2 speech in terms of four features (i.e., intonation, rhythm, word stress, and overall proficiency), the evaluators could listen to the whole recording of the

diagnostic passage or three sentences cut from the recording for each feature. Five-point scales were utilized in order to evaluate the L2 speech production. The evaluators were not limited in the number of times they could listen to the speech samples. In order to control any effects due to the order of presentation, the stimuli were presented in a different order to half of the evaluators in each group. Both groups of evaluators were asked to focus their attention more on suprasegmental characteristics than on segmental sounds. The inter-rater reliability of the English speaking evaluators reached .96, and that of the Korean evaluators was .78.[2] It could be assumed that the evaluators were consistent in assessing L2 speech.

4. Results

The important question in this study is whether or not there are any differences in the assessments of native English speaking evaluators and nonnative English speaking evaluators in terms of suprasegmental features and overall proficiency. Another question investigated is whether any such discrepancies are shown more clearly for L2 speakers with a lower proficiency level than for those with a higher proficiency level. Finally, this study examines the question of whether there are any differences between the four pronunciation features.

Assessment scores for three suprasegmental features and overall fluency were coded for statistical analysis. The box plots in Figures 1 and 2 display the distribution of the mean scores of the two proficiency groups provided by two groups of evaluators in regard to four pronunciation features.[3] The mean scores for each feature are the average of the scores given by the eight evaluators in each group. Figure 1 presents the mean scores of ten speakers in the lower proficiency level group, whereas Figure 2 shows the mean scores of another ten speakers in the higher proficiency level group.

---

[2] Inter-rater reliability is the degree of agreement among raters. In this paper Cronbach's alpha reliability coefficients were computed using the SPSS program (version 18.0) as an index of inter-rater consistency. According to Portney and Watkins (2009), coefficients below .50 represent poor reliability, coefficients from .50 to .75 indicate moderate reliability, and values above .75 suggest good reliability.
[3] The box plots display the distribution of data based on the five number summary: minimum, first quartile, median, third quartile, and maximum.
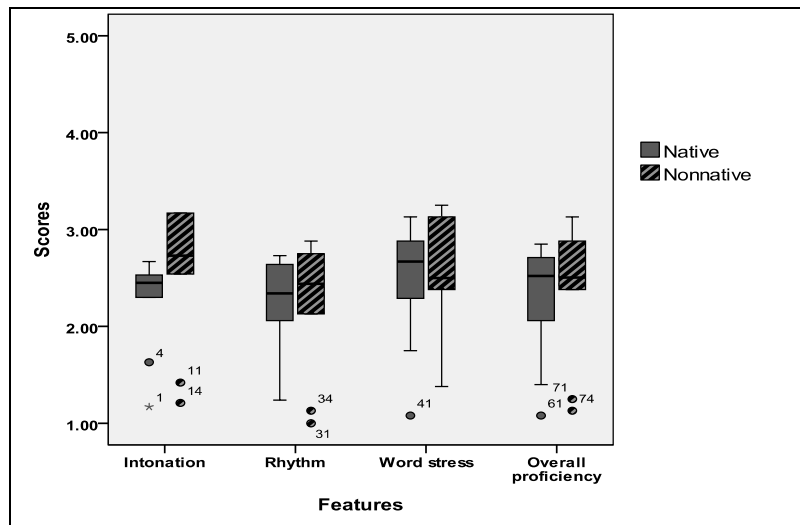
**Figure 1. Mean scores of the lower proficiency level group
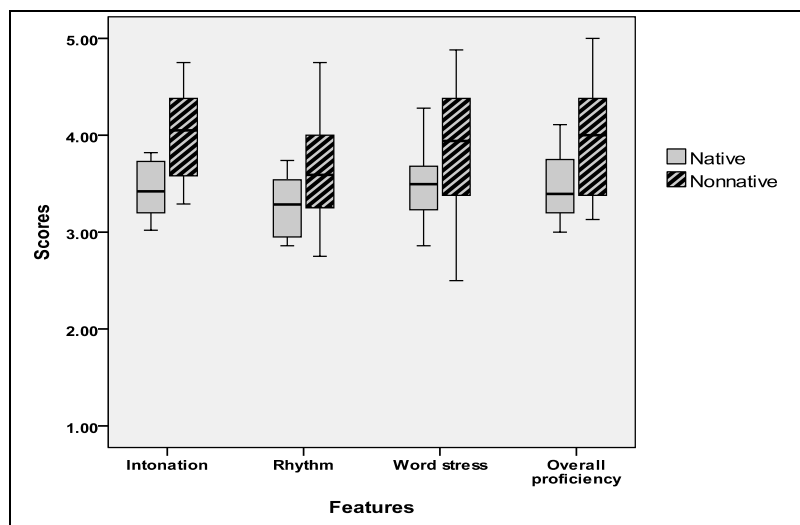rated by native and nonnative evaluators**

**Figure 2. Mean scores of the higher proficiency level group
rated by native and nonnative evaluators**

As can be seen in Figures 1 and 2, nonnative English speaking evaluators provided higher scores than native English speaking evaluators for L2 speech from both the lower and higher proficiency level groups. In other words, the nonnative evaluators' ratings of suprasegmental features were stricter than that of native evaluators. Furthermore, the discrepancies in the assessments between native and nonnative evaluators were more clearly shown for the higher proficiency level group than for the lower one. Moreover, for the higher proficiency level group, there was substantially more variation in the assessments by nonnative evaluators, which range from 2.5 to 4.88, whereas the assessments by native evaluators range from 2.86 to 4.28.

The multi-way analysis of variance (ANOVA) was conducted on the mean rating scores in order to analyze the effects of nativeness of evaluators (native English speaking evaluators vs. advanced-level nonnative English speaking evaluators), L2 proficiency (higher proficiency L2 speakers vs. lower proficiency L2 speakers), and pronunciation features (intonation, rhythm, word stress, and overall proficiency). The results of the statistical analysis are shown in Table 1.

Table 1. Multi-way ANOVA of mean scores by native and nonnative evaluators

Source	Type III Sum of Squares	df	Mean Square	F	*Sig.*
Nativeness	3.153	1	3.153	9.559	.002*
L2 proficiency	65.127	1	65.127	197.459	.000*
Pronunciation features	1.457	3	.486	1.472	.225

* p < 0.05

Overall, there was a significant evaluators' nativeness effect [$F(1, 14)=9.559$, $p=.002$] with the native English speaking evaluators providing lower rating scores than the nonnative evaluators. There was also a significant effect of L2 proficiency [$F(1, 18)=197.459$, $p=.000$], yielding much lower rating scores for the group at a lower L2 proficiency level than for the one at a higher L2 proficiency level. There was no significant effect in pronunciation features. Differences in the rating scores between the four pronunciation features were not statistically significant [$F(3, 16)=1.472$, $p=.225$]. Also, there was no significant effect in interaction between the factors. Thus, although there were significant main effects of nativeness, the nativeness effect was shown in relation to neither L2 proficiency levels nor pronunciation features.

As mentioned earlier, some differences were revealed between the higher proficiency level group and the lower proficiency level one. The discrepancies in the assessments between the native and nonnative

evaluators were more clearly shown in the higher proficiency level group. The mean ratings obtained from native and nonnative evaluators for each speaker group were submitted to *t*-tests. Table 2 presents statistical differences in the mean rating scores between the two groups of evaluators for each speaker group separately.

Table 2. *T*-tests for the difference of native and nonnative evaluators' mean scores for each speaker group

	Native evaluators		Nonnative evaluators			
	Mean	S.D.	Mean	S.D.	*t*	*Sig.* (two-tailed)
Lower proficiency	2.302	.542	2.426	.672	-1.069	.288
Higher proficiency	3.421	.352	3.859	.632	-5.028	.000*

* p < 0.05

The t-tests indicate that significant differences in the assessments between the two groups of evaluators were shown only for the higher proficiency level speakers (*t*=-5.028, *p*=.000). For the lower proficiency level group, there were no significant differences in the assessment scores between native and nonnative evaluators (*t*=-1.069, *p*=.288). The awkwardness of suprasegmental features in the pronunciation of the higher proficiency level group was less salient to the nonnative evaluators than to the native evaluators.

Next, in order to compare the higher proficiency group rating scores of the native and nonnative evaluators in terms of four features, a series of independent *t*-tests were implemented. Table 3 indicates differences in the mean rating scores between the two groups of evaluators.

Table 3. *T*-tests for the difference of native and nonnative evaluators' mean scores for the high proficiency group

Feature	Mean difference (native - nonnative)	*t*	*Sig.* (two-tailed)
Intonation	-.548	-2.892	.010*
Rhythm	-.360	-1.529	.144
Word stress	-.336	-1.286	.215
Overall proficiency	-.507	-2.157	.045*

* p < 0.05

As shown in Table 3, for the higher proficiency group, the *t*-tests yielded a significant difference between the two groups of evaluators for intonation (*t*=-2.892, *p*=.010), and a marginally significant difference for overall proficiency (*t*=-2.157, *p*=.045). Nonnative evaluators gave significantly higher ratings than native evaluators for intonation and overall proficiency. There were no statistically significant differences between the two groups of evaluators for the other features. That is, although the nonnative ratings were also slightly higher than the native ratings in terms of the other features, the difference was not significant (*t*=-1.529, *p*=.144 for rhythm; *t*=-2.157, *p*=.215 for word stress). Such findings indicate that the aspects of foreign accent related to suprasegmental features and overall proficiency is less noticeable to nonnative evaluators than to native evaluators. These results are also consistent with previous findings (Lee 2010, Kang and Rhee 2011, Sung 2011). However, it should be noted that the discrepancies in the rating scores between native and nonnative evaluators were shown only in the high proficiency group. The nonnative evaluators were able to recognize poor production of suprasegmental features shown by the low proficiency level L2 speakers. The nonnative evaluators, however, were less sensitive to awkwardness in intonation and stress shown by L2 speakers of high proficiency level than were the native evaluators.

When examining the rating scores for individual speakers in the higher proficiency level group, the differences between native and nonnative assessments were most clearly shown for three speakers. There was no overlap in the scores of these individual speakers given by native and nonnative evaluators. It seems that the nonnative evaluators' assessments sometimes depended more on the overall proficiency of the speakers than on the suprasegmental features of speech.

In the comments section of the assessment form, the native evaluators indicated that overall, the L2 speech samples sounded flat and unnatural. Many native evaluators pointed out the strangeness of the intonation at the end of sentences. The intonation patterns of some speakers were noted as being overly dramatic due to the abrupt sentence-final falls in pitch. In contrast, in some samples the falling intonation was too slight and gradual. In particular, the falling intonation at the end of the statements was often not discernible enough to indicate completion. Intonation was sometimes significantly changed at certain points where some speakers attempted to self correct their pronunciation.

The speech rate was another major problem, especially for the higher proficiency level speakers. Many of the higher proficiency level speakers read the passage too quickly even though they were asked to read it at a natural rate during the production experiment. This fast speech made it difficult to provide appropriate pitch patterns. Sometimes numerous hesitations were also evident in fast speech, and these unexpected breakdowns interrupted the natural flow of speech. Moreover, the L2 speakers had some degree of trouble with the location of pauses in long

sentences, which made their speech difficult to understand.

Furthermore, the Korean speakers had problems with lexical stress and rhythm. The native evaluators commented that many speakers at both proficiency levels enunciated the vowel sounds of each syllable too precisely, without reduction in unstressed syllables. An awkward rising tone was also shown at the end of multisyllabic words such as "intonation" or "pronunciation". Also, the speakers often stressed too many words, yet also occasionally neglected to stress important words. The insertion of an extra vowel at the end of a word was frequently indicated, too.

The nonnative evaluators commented on inappropriate intonation and stress patterns for the lower proficiency level speakers just as the native evaluators did. However, for the higher proficiency level speakers, the nonnative evaluators' comments and ratings were quite different from those of the native evaluators. For one higher proficiency level speaker, many native evaluators indicated that her speech lacked appropriate pitch variation because of overemphasis of certain syllables, and that her statement-final falling intonation was odd and inconsistent. However, several nonnative evaluators commented that her intonation patterns were fairly accurate, although some sentences sounded somewhat monotonous.

5. Discussion and Conclusion

The main purpose of this study is to compare native and nonnative evaluators' assessments of suprasegmental features and overall proficiency of L2 speech. Assessments by native and nonnative evaluators of the L2 speech of speakers at two different proficiency levels are also compared, and an attempt is made to determine specifically which suprasegmental features most clearly reflect the differences in assessments between the two groups of evaluators.

In order to investigate the differences in assessment between the native and nonnative evaluators, two groups of evaluators were employed, consisting of eight native English speaking evaluators and eight advanced-level nonnative English speaking evaluators, respectively. Both groups of evaluators rated speech samples from two groups of Korean learners of English, with one consisting of ten lower proficiency level speakers and the other, of ten higher proficiency level speakers. The three suprasegmental features such as intonation, rhythm, and word stress, as well as overall proficiency, were rated on a five-point scale.

The results based on the mean rating scores show that there were significant differences in the assessments between native and nonnative evaluators. The scores given by the native evaluators turned out to be significantly lower than those of nonnative evaluators. Furthermore, the differences in the assessments between native and nonnative evaluators varied between the two groups of L2 speakers. When comparing the assessment scores of the two groups of evaluators for the two groups of

speakers separately, the statistical analyses revealed that the discrepancies in the assessments between the two groups of evaluators were shown only for the higher proficiency group. That is, for the lower proficiency group, the rating scores were consistent between native and nonnative evaluators. However, for the higher proficiency group, the native evaluators' rating scores were significantly lower than those of nonnative evaluators. It is evident that the Korean evaluators were able to notice completely inappropriate pitch patterns such as a raised pitch at the end of a wh-question sentence. However, slightly inappropriate rising or falling pitch patterns were not salient to the nonnative evaluators.

Many previous studies have shown that in the perception of foreign accents, suprasegmental features are more salient than, or at least as salient as, individual segments (Anderson-Hsieh et al. 1992, Derwing and Munro 1997, Derwing and Rossiter 2002, Hahn 2004, Boula de Mareuil and Veru-Dimulescu 2006). Thus, L2 learners of English should pay more attention to suprasegmental factors speech. Furthermore, previous research (Elliott 1977, Derwing et al. 1998, Moyer 1999) has revealed that prosodic instruction could result in significant improvements in L2 speech proficiency. Therefore, there are pedagogical reasons to focus on suprasegmental aspects of L2 speech. In the present study, the advanced-level nonnative evaluators and the nonnative speakers in the higher proficiency group who produced speech samples were all graduate students majoring in English Education, and most of them expressed the intention to be English instructors at public or private schools in the future. If L2 instructors cannot produce appropriate intonation and stress patterns or perceive the irregularities of prosodic features of L2 speech, it will clearly be more difficult for them to teach such prosodic essentials to their students.

The discrepancies in the assessments between native and nonnative evaluators found in the results of the present study support the results of previous research (Lee 2010, Kang and Lee 2011, Sung 2011). However, the present findings were not in exact accordance with the results shown in Rhee and Park (2003), where very high correlations were found between native and nonnative evaluators. It must be noted that there were some differences in between Rhee and Park (2003)'s study and the present study. The former examined just the general proficiency of L2 speakers, and the samples of L2 speech were produced by only one group of L2 speakers at various levels of English proficiency. The present study, however, focuses on suprasegmental features of L2 speech, and the speech samples were obtained from two groups of L2 speakers divided according to proficiency level.

For the higher proficiency group, the average ratings between native and nonnative evaluators were compared with regard to four features. The results indicate that nonnative evaluators provided significantly higher scores than native evaluators for intonation and overall proficiency. The

awkwardness of intonation patterns of L2 learners was not salient to nonnative evaluators even though the nonnative evaluators were all advanced-level L2 learners. This phenomenon was also shown in the comments sections of the assessment forms. For example, many native evaluators pointed out that pitch variations were too abrupt or indiscernible at the end of the statements. However, most of the nonnative evaluators commented that the intonation patterns were correct aside from being a little flat. In order to increase nonnative evaluators' sensitivity to pitch variations, suprasegmental aspects of L2 speech should receive more attention in instructional settings. Considering that intonation plays a key role in the perception of foreign accent (Pickering 1999, Jilka 2000), the acquisition of appropriate intonation patterns is an essential phase of L2 learning.

The present results showing higher ratings given by nonnative evaluators than those by native evaluators could have been affected by the shared native language background of the L2 speakers and L2 evaluators. Previous research (van Wijngaarden 2001, Bent and Bradlow 2003, Jongman and Wade 2007) has shown that for nonnative evaluators, the intelligibility of nonnative speakers who had the same native language was greater than or equal to the intelligibility of native speakers. For future studies, it may be interesting to compare the assessments of L2 suprasegmentals between native and nonnative evaluators when nonnative evaluators and nonnative speakers do not share the same native language. In other words, an empirical question which might deserve to be examined in the future is whether or not discrepancies in assessments of L2 speech suprasegmentals between native and nonnative evaluators will decrease when the nonnative evaluators judge the L2 speech of speakers with a different kind of foreign accent.

Another interesting topic for future study might be to investigate whether overt training in L2 suprasegmental features has any effect on the development of L2 speech proficiency for different groups of speakers divided according to native languages with different prosodic structures. Moreover, future studies may explore the relationship between perception and production development of L2 speech through the direct instruction of L2 suprasegmental features.

Appendix A. Diagnostic passage[4]

Are you a native speaker of English? If you are not, your pronunciation and intonation may reveal your country of origin. Why do learners find it so difficult to speak a foreign language without an accent? There are a few different reasons. The first important factor in learning pronunciation is age.

---

[4] This passage is a modified version of the one found in a course book (Celce-Murcia et al. 2010: 481).

Young children have the ability to pronounce a second language just like a native speaker, while most older learners usually find it difficult to do so.

What is another factor that influences the way you pronounce a foreign language? Your first language. For example, most English speakers can identify people from France just by hearing their French accent. They can also recognize Arabic or Spanish speakers when talking on the telephone, merely by listening with open ears. Does this mean that you can never change your accent? Of course not! But if you want to change your pronunciation, you will really have to work at it. Through a combination of careful listening, strong ambition and hard work, some day you may be able to sound like a native speaker.

What else do you need? Effective practice strategies, accurate information about the sounds of English, and lots of patience. Will you give up? Or will you progress? It's all up to you. If you are ready to do what it takes, you can improve. Good luck, and get to work!

Appendix B. Assessment form[5]

1) Listen to the entire passage and read the directions.
2) Listen to the sample portions.
3) Rate the speaker on a scale from 1 (strong foreign accent) to 5 (native-like accent).
4) Provide comments.

1. Intonation

1) yes/no questions

Listen for the following **yes/no questions** in the recording, and rate the intonation.

Are you a native speaker of English?
Does this mean that you can never change your accent?
Will you give up?

Strong foreign accent   Native-like accent
        1   2   3   4   5
    _____   _____   _____   _____   _____

Comments: _____

---

[5]   This is a modified version of one found in a course book (Celce-Murcia et al. 2010: 487).

2) wh-questions

Listen for the following **wh-questions** in the recording, and rate the intonation.
Why do learners find it so difficult to speak a foreign language without an accent?
What is another factor that influences the way you pronounce a foreign language?
What else do you need?

Strong foreign accent   Native-like accent
         1   2   3   4   5

    _____   _____   _____   _____   _____

Comments: _____

3) statements

Listen to her **statements**. Does her voice pitch fall at the end of thoughts and statements to show completion?

If you are not, your pronunciation and intonation may reveal your country of origin.
There are a few different reasons.
The first important factor in learning pronunciation is age.

Strong foreign accent   Native-like accent
         1   2   3   4   5

    _____   _____   _____   _____   _____

Comments: _____

2. Sentence stress and rhythm

Consider the following sentences. How was her overall rhythm? Did she use the appropriate stress patterns?

Why do learners find it so difficult to speak a foreign language without an accent?
There are a few different reasons.
The first important factor in learning pronunciation is age.
Young children have the ability to pronounce a second language just like a native speaker, while most older learners usually find it difficult to do so.

Strong foreign accent   Native-like accent
          1   2   3   4   5
     _____   _____   _____   _____   _____

Comments: _____

3. Word stress

Consider the following underlined words in the sentences. Did she stress the appropriate syllables?

Are you a underline native speaker underline of underline English underline?
If you are not, your underline pronunciation underline and underline intonation underline may underline reveal underline your underline country underline of underline origin underline.
Why do underline learners underline find it so underline difficult underline to speak a underline foreign underline language underline without an underline accent underline?

Strong foreign accent   Native-like accent
          1   2   3   4   5
     _____   _____   _____   _____   _____

Comments: _____

4. Overall proficiency/fluency

Rate the overall proficiency or fluency of her speech.

Strong foreign accent   Native-like accent
          1   2   3   4   5
     _____   _____   _____   _____   _____

Comments: _____

REFERENCES

ANDERSON-HSIEH, JANET, RUTH JOHNSON, and KENNETH KOEHLER. 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, JANET and HORABIL VENKATAGIRI. 1994. Syllable duration and pausing in the speech of Chinese ESL speakers. *TESOL Quarterly* 28, 807–812.

BENT, TESSA and ANN R. BRADLOW. 2003. The interlanguage speech intelligibility benefit. *Journal of the Acoustical Society of America* 114.3, 1600-1610.

BOND, ZINNY. 1999. *Slips of the ear: Errors in the perception of casual conversation*. San Diego, CA: Academic Press.

BOULA DE MAREUIL, PHILIPPE and BIANCA VIERU-DIMULESCU. 2006. The contribution of prosody to the perception of foreign accent. *Phonetica* 63, 247-267.

CELCE-MURCIA, MARIANNE, DONNA M. BRINTON , JANET M. GOODWIN, and BARRY GRINER. 2010. *Teaching Pronunciation: A course book and reference guide.* Cambridge University Press.

CUCCHIARINI CATIA, HELMER STRIK, and LOU BOVES 2000. Different Aspects of Expert Pronunciation Quality Ratings and Their Relation to Scores Produced by Speech Recognition Algorithms. *Speech Communication* 30, 109-119.

DERWING, TRACEY M. and MURRAY J. MUNRO. 1997. Accent, intelligibility, and comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition* 19, 1–16.

DERWING, TRACEY M., MURRAY J. MUNRO, and GRACE E. WIEBE. 1998. Evidence in favor of a broad framework for pronunciation instruction. *Language Learning* 48, 393-410.

DERWING, TRACEY M. and MURRAY J. MUNRO. 2001. What speaking rates do non-native listeners prefer? *Applied Linguistics* 22, 324–227.

DERWING, TRACEY M. and MARIAN J. ROSSITER. 2002. ESL learners' perceptions of their pronunciation needs and strategies. *System* 30, 155-166.

ELLIOTT, RAYMOND A. 1997. On the teaching and acquisition of pronunciation within a communicative approach. *Hispania* 80, 95-108.

FIELD, JOHN. 2005. Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly* 39.3, 399–423.

FLEGE, JAMES. E. 1984. The detection of French accent by American listeners. *Journal of the Acoustical Society of America* 76, 692-707.

_____. 1988. Factors affecting degree of perceived foreign accent in English sentences. *Journal of the Acoustical Society of America* 84, 70-79.

HAHN, LAURA D. 2004. Primary stress and intelligibility: Research to motivate the teaching of suprasegmentals. *TESOL Quarterly* 38.2, 201-223.

JILKA, MATTHIAS. 2000. *The Contribution of Intonation to the Perception of Foreign Accent.* PhD Dissertation. Institute of Natural Language Processing. University of Stuttgart.

JONGMAN, ALLARD and TRAVIS WADE. 2007. Acoustic variability and perceptual learning: The case of non-native accented speech. In O.S. Bohn and M. J. Munro (ed.). *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, 135-150.

KANG, SEOKHAN 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.

LEE, BORIM, SUSAN G. GUION, and TETSUO HARADA. 2006. Acoustic analysis of the production of unstressed English vowels by early and late Korean and Japanese bilinguals. *Studies in Second Language Acquisition* 28, 487–513.

LEE, CHOON-HEE. 2010. Improving inter-rater reliability in oral proficiency test at college level. *Modern English Language and Literature* 54.1, 367-385.

MOYER, ALENE. 1999. Ultimate attainment in L2 phonology. The critical factors of age, motivation, and instruction. *Studies in Second Language Acquisition* 21.1, 81-108.

MUNRO, MURRAY J. and TRACEY M. DERWING. 1995. Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech* 38, 289-306.

PICKERING, LUCY. 1999. *An analysis of prosodic systems in the classroom discourse of native speaker and nonnative speaker teaching assistants*. PhD Dissertation. Gainesville, University of Florida.

PORTNEY, LESLIE G. and MARY P. WATKINS. 2009. *Foundations of Clinical Research: Applications to Practice*. Upper Saddle River, NJ: Pearson Prentice Hall.

RHEE, SEOK-CHAE and JEON-KYOO PARK. 2003. The correlation study on evaluation of Korean-spoken English for Korean/native English/ITL. *Proceedings of Korean Phonetics Society*, 83-87.

SUNG, EUNKYUNG. 2011. L2 speech assessment by native and nonnative listeners: Segmental sounds and suprasegmental features. *Studies in Phonetics, Phonology and Morphology* 17.1, 25 - 41.

TAJIMA, KEIICHI, ROBERT PORT, and JONATHAN DALBY. 1997. Effects of speech timing on intelligibility of foreign accented English. *Journal of Phonetics* 25, 1-24.

THOMPSON, IRENE. 1991. Foreign accents revisited: The English pronunciation of Russian immigrants. *Language Learning* 41.2, 177-204.

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, 1–30

VAN WIJNGAARDEN, SANDER J. 2001. Intelligibility of native and non-native Dutch speech. *Speech Communication* 35, 103-113.

YUN, WEONHEE. 2009. An analysis of the Korean inter-rater difference in evaluating English pronunciations of Korean speakers. *Studies in Foreign Language Education* 23.2, 85-104.

Eunkyung Sung
Department of English
Cyber Hankuk University of Foreign Studies
270 Imun-dong, Dongdaemun-gu,
Seoul, Korea 130-791
e-mail: eks@cufs.ac.kr