

L2 speech assessment by native and nonnative listeners: segmental sounds and suprasegmental features*

Eunkyung Sung
(Cyber Hankuk University of Foreign Studies)

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. This study compares native and nonnative listeners' assessments of L2 speech produced by Korean learners of English in terms of six pronunciation features: intonation, stress, reduced speech, linking, individual sounds, and consonant clusters. Speech samples from fifteen advanced English learners were recorded and then rated by native and nonnative listeners. The results show that nonnative listeners provide higher ratings for L2 speech than native listeners for all of the features except individual sounds. For three suprasegmental features – intonation, stress, and reduced speech – nonnative listeners' ratings are significantly higher than those of native listeners, indicating that suprasegmental features are less salient to nonnative listeners than to native listeners. The native listeners also point out that overall, the L2 speech samples sound flat and unnatural due to the lack of appropriate intonation and stress. Considering that in L2 speech pronunciation the role of prosodic or suprasegmental features outweighs that of segmental sounds, the focus on suprasegmental features needs to be increased in L2 pronunciation instruction. (Cyber Hankuk University of Foreign Studies)

Keywords: L2 speech, assessment, native listeners, nonnative listeners, segmental sounds, suprasegmental features

1. Introduction

In second language (L2) learning and teaching, the importance of pronunciation for successful communication has been gaining recognition. It has also been recognized that not only segmental but also suprasegmental errors can impede communication between native speakers and L2 learners (Anderson-Hsieh et al. 1992, Derwing and Rossiter 2002, Hahn 2004, Boula de Mareuil and Veru-Dimulescu 2006).

Although pronunciation has been getting more attention in L2 acquisition research, the assessment of L2 pronunciation has been relatively neglected. Accented L2 speech is usually assessed holistically with the help of native listeners. Much less is known about the assessment of accented L2 speech by L2 listeners. Furthermore, there has been little research which contrasts the assessment of native English listeners with that of nonnative listeners.

* This work was supported by the Cyber Hankuk University of Foreign Studies Research Fund of 2011. I am grateful to three anonymous reviewers for their insightful comments and suggestions on this paper.

This study compares assessments of native listeners and those of nonnative listeners with regard to how they perceive Korean speakers' English pronunciation. Fifteen Korean graduate students' readings of an English passage were recorded and assessed by two native English listeners and two nonnative listeners; in the nonnative group, one assessment was a self-assessment, and the other was done by a peer. Six pronunciation features were chosen to be rated on a five-point scale – intonation, stress, reduced speech, linking, individual sounds, and consonant clusters. These features are considered essential components of pronunciation that are assessable. Out of these six features, intonation, stress, reduced speech, and linking are suprasegmental properties. The following research questions will be investigated in this study: 1) Do native English-speaking listeners and advanced-level nonnative listeners assess L2 speech differently? 2) If so, for which features of L2 pronunciation – segmental sounds or suprasegmental features - do assessments most widely vary?

2. Previous Studies

2.1 Segmental sounds and suprasegmental features in L2 pronunciation

Recent research has shown that suprasegmental or prosodic features contribute more than, or at least equally to, segmental sounds to the perception of a foreign accent in L2 speech by native listeners (Anderson-Hsieh et al. 1992, Hahn 2004, Boula de Mareuil and Veru-Dimulescu 2006). With some exceptions¹, recent research on pronunciation pedagogy has focused on elements of suprasegmentals and prosody. Anderson-Hsieh et al. (1992) provide evidence of the importance of prosodic variables in the comprehensibility of oral production. They found that judgments of comprehensibility of speech samples of nonnative speakers are closely related to independent measures of overall prosody. Hahn (2004) indicates that suprasegmentally dissimilar but segmentally nearly identical speeches are evaluated and comprehended differently by native English speakers. Hahn also mentions that incorrect placement or complete absence of suprasegmental features can lead to a poor assessment of the speaker. In other words, native speakers tend to evaluate nonnative speakers more favorably when their speech maintains appropriate primary stress patterns. Boula de Mareuil and Veru-Dimulescu (2006) demonstrate that prosody (timing and melody) outweighs segmental cues in the perception of Spanish-accented Italian and Italian-accented Spanish, by using a prosodic transplantation paradigm.

In a comparison between segmental and suprasegmental approaches to

¹ Jenkins (2000; 2002) denies the importance of suprasegmental usage in communication among nonnative speakers, and proposes a pronunciation syllabus which advocates a segmental-centered approach to pronunciation instruction.

pronunciation instruction, Elliott (1997) shows that the teaching of segmental sounds does not result in significant improvements in spontaneously produced speech. Derwing et al. (1998) also illustrate that only suprasegmental teaching improves comprehensibility in spontaneous speech. Although segmental instruction is beneficial when students are asked to read sentences aloud, the effects do not appear to transfer to spontaneously produced utterances. Pennington and Ellis (2000) demonstrate that learners who receive explicit instruction on stress can improve their stress production significantly. Derwing and Rossiter (2002) point out that many pronunciation textbooks have focused on L1-L2 segmental differences, as well as discrimination and articulation of individual segments in single-syllable words. Course books consist primarily of phonetic symbols, charts, and diagrams of the place and manner of sound articulation. Further, lists of minimal pairs for practicing individual sounds appear more often than tasks for learning prosodic elements. They indicate that although a shift of attention toward the prosodic aspects of pronunciation such as intonation, stress, and rhythm has led to calls for a broader approach to speech performance, course books do not include enough practice exercises for those prosodic features. Derwing and Rossiter (2003) also show that pronunciation instruction for suprasegmentals increases the intelligibility of nonnative speakers of English.

Previous research manifests that although researchers and educators have called for more attention to the incorporation of suprasegmental elements in pronunciation teaching, practical methods to teach and practice suprasegmentals need to be developed.

2.2 Assessment by native and nonnative listeners

For assessing the intelligibility of L2 speech, previous studies have found that the speech of nonnative speakers is more intelligible to nonnative listeners than to native listeners (van Wijngaarden 2001, Bent and Bradlow 2003, Jongman and Wade 2007). It has also been shown that speech uttered by a high-proficiency nonnative speaker from the same native language background is as intelligible to the listener as speech from a native speaker. Furthermore, for nonnative listeners, the intelligibility of a nonnative speaker from a different native language background is greater than the intelligibility of a native English speaker, giving rise to “the mismatched interlanguage speech intelligibility benefit” (Bent and Bradlow 2003).

When nonnative listeners rate nonnative speakers’ foreign accents, the raters’ own linguistic background affects the reliability of the assessment. Flege (1984) and Thompson (1991) have shown that linguistically experienced listeners are more reliable than inexperienced judges in estimating foreign language learners’ speech intelligibility. Flege (1988) and MacKay et al. (2006) argue that there is similarity between native and

nonnative judges when assessing foreign accents, based on the lack of statistical differences between foreign accent ratings by these two groups of listeners.

Most previous studies on the perceptions and assessments of native and nonnative listeners have focused on the holistic intelligibility or comprehensibility of L2 speech. Little attention, however, has been given to comparisons of L2 speech assessment between native and nonnative listeners in terms of segmental and suprasegmental features.

In Korea, nonnative instructors often teach English in formal instructional settings, and they also assess learners' pronunciation. The goal for the present study is to examine differences between these two types of listeners, native and advanced nonnative listeners, with regard to how they assess Korean speakers' English pronunciation. This study focuses on six pronunciation features: intonation, stress, sound reduction, linking, individual sounds and consonant clusters.

3. Methods

3.1 Participants

Participants in this study were comprised of 15 female Korean graduate students at H University in Seoul and two native English speakers who were English instructors at the same university. The Korean learners of English ranged in age from 24 to 37 years old and all majored in English education at the graduate school. Their English proficiency was high based on TOEIC scores ranging from 860 to 970 (the mean score was 923). They were enrolled in the English pronunciation course and did not seem to have any difficulty understanding English lectures or communicating with their peers or the instructor in English. Therefore, it can be said that they were all advanced learners of English. These Korean learners of English participated in the production and evaluation of their own speech tasks as well as those of their peers. Two native English speakers evaluated the 15 Korean speakers' speech production independently of each other. Thus, two native English speakers and two nonnative English speakers (self and a peer) rated each Korean speaker's pronunciation.

3.2 Materials

3.2.1 Diagnostic passage

A diagnostic passage was used to assess the participants' command of pronunciation features. The diagnostic passage, a modified version of the one found in a course book (Celce-Murcia et al. 2010: 481), was designed to assess the learner's ability to produce certain segments and consonant clusters as well as their ability to use appropriate stress and intonation. The

passage was composed of 18 sentences including 6 interrogative sentences (see Appendix A). Although a reading of the passage did not represent the learner's spontaneous speech, it allowed the evaluators to obtain a sampling of individual errors.

3.2.2 Assessment form

An assessment form, which was also a slightly modified version of one found in a course book (Celce-Murcia et al. 2010: 487), was used to assess the participants' recordings. The assessment form consisted of six features and eleven sub-categories: intonation (yes/no questions, wh-questions, statements), stress (sentence prominence, word stress), reduction (reduction of function words, vowel reduction), linking, individual sounds (vowels, consonants), and consonant clusters. Each sub-category item was measured on a five-level scale, with 5 representing native-like pronunciation and 1 indicative of a strong foreign accent. There was also a field for comments below the five-level scale measurement for each item (see Appendix B).

3.3 Procedure

The participants were asked to read the diagnostic passage at a natural rate. They were allowed time to practice before recording the passage to be sure that they knew every word in the passage. Their speech production was recorded using a DAT recorder (Tascam US-122 Audio Interface) and a microphone (Audio Technica ATM75) in a quiet room. After the recording, the individual participants listened to and rated their own speech production as well as a peer's. The two native English speakers rated the Korean participants' pronunciation independently of each other. Thus, there were two types of assessment: the native assessment and non-native assessment. Five-point scales were utilized in order to assess the participants' speech production.

4. Results

Evaluation scores for six pronunciation features and eleven sub-categories were coded for statistical analyses. Table 1 depicts descriptive statistics for the scores provided by the two groups of listeners, including the means and standard deviations. The mean score for each feature is the average of the two scores given by the two listeners in each group. The native mean score is the average of the two scores provided by the two native English listeners, and the nonnative mean score is the average of the self-rating and the peer-rating scores. For all features the scores provided by native listeners were lower than those of nonnative listeners except for one feature: individual sounds. The nonnative listeners' rating for consonant

production was stricter than that of native listeners.

Table 1. Mean rating scores by native and nonnative listeners

Feature	Sub-category	Rater	Mean Score	SD
Intonation	Yes/No questions	Native	3.3333	.6172
		Nonnative	3.7667	.4169
	Wh-questions	Native	3.2000	.5916
		Nonnative	3.4000	.6866
	Statements	Native	3.3000	.5278
		Nonnative	3.8000	.6492
Stress	Prominence	Native	3.0000	.6268
		Nonnative	3.6333	.5164
	Word stress	Native	3.0000	.7319
		Nonnative	3.5000	.5976
Reduced speech	Reduction of function words	Native	2.8333	.5233
		Nonnative	3.2333	.6510
	Reduction of vowels	Native	3.2333	.5300
		Nonnative	3.5000	.4629
Linking		Native	3.3333	.6455
		Nonnative	3.5333	.8121
Sounds	Vowels	Native	3.3667	.6399
		Nonnative	3.4667	.6114
	Consonants	Native	3.4333	.7528
		Nonnative	3.1333	.6935
Consonant clusters		Native	3.3333	.6986
		Nonnative	3.6333	.8958
Total		Native	3.2294	.5832
		Nonnative	3.5091	.6456

Some discrepancies are shown between the native and nonnative listeners' assessments of English pronunciation of the Korean speakers. Figure 1 displays a comparison of mean scores given by native and nonnative listeners with regard to six pronunciation features. The mean score of each feature is the average of the mean scores of sub-categories. For example, the mean score of intonation is the average of the mean scores of three sub-categories: yes/no questions, wh-questions, and statements.

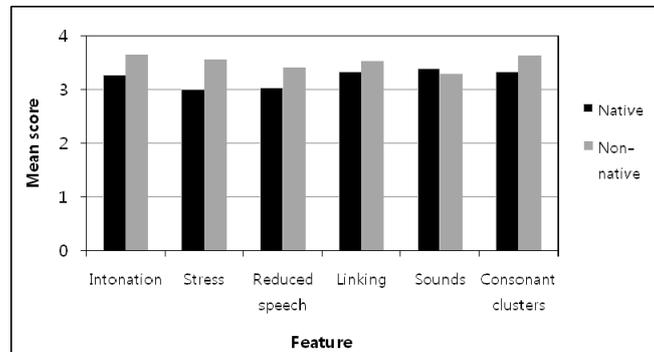


Figure 1. Mean scores by native and nonnative listeners

Among the six features, nonnative listeners' average rating scores were higher than those of native listeners for all the features except one: individual sounds. The native listeners' average rating score was higher than that of nonnative listeners for individual sounds. In other words, the native listeners were less strict in their assessment of segmental sounds than the nonnative listeners. For the other features, the native listeners presented stricter rating than the nonnative listeners.

The two-way analysis of variance (ANOVA) was conducted on the mean rating scores, one with Group (native or nonnative listeners) as a between-subject factor and the other with Feature (intonation, stress, reduced speech, linking, sounds, or consonant clusters) as a within-subject factor. The results of the statistical analysis are shown in Table 2.

Table 2. Two-way ANOVA of mean scores by native and nonnative listeners

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Group	3.520	1	3.520	9.293	.003*
Feature	1.884	5	.377	.995	.422
Group * Feature	1.844	5	.369	.974	.436

* $p < 0.05$

Overall, there was a significant Group effect [$F(1, 28) = 9.293, p = .003$] with the native listeners providing higher rating scores than the nonnative listeners. However, there was no significant effect in both Feature and the interaction of Group and Feature. In order to consider the rating on each feature relative to the other features, a repeated measures ANOVA was carried out separately on each listener group. The results revealed that there was no significant difference between the six features for both

listener groups [$F(14, 75) = 2.070$, $p = .079$ for native listeners; $F(14, 75) = 2.075$, $p = .079$ for nonnative listeners].

Next, in order to compare the rating scores between the native and nonnative listeners in terms of six features, a series of independent t-tests were implemented. Table 3 indicates differences in the mean rating scores between the two listener groups.

Table 3. T-tests for the difference of mean scores by native and nonnative listeners

Feature	Mean difference (native - nonnative)	t	Sig. (two-tailed)
Intonation	-.3780	-2.133	.042*
Stress	-.5667	-3.301	.005*
Reduced speech	-.3833	-2.279	.030*
Linking	-.2000	-.747	.461
Sounds	.1000	.453	.654
Consonant clusters	-.3000	-1.023	.315

* $p < 0.05$

The t-tests yielded a significant difference between the two groups for intonation, stress, sound reduction ($t = -2.133$, $p = .042$ for intonation; $t = -3.012$, $p = .005$ for stress; $t = -2.279$, $p = .030$ for reduced speech). Nonnative listeners gave significantly higher ratings than native listeners for three features: intonation, stress, and reduced speech. There were no statistically significant differences between the two groups of listeners for the other features. The nonnative rating was slightly lower than the native rating only in segmental sounds, but the difference was not significant ($t = .453$, $p = .654$). Such findings indicate that the foreign accent related to suprasegmental features is less noticeable to nonnative listeners than to native listeners.

In the comments section of the assessment form, the native listeners pointed out that overall, the L2 speech samples sounded flat, without major fluctuations in intonation and stress. The native listeners also found that the intonation contour was awkward and somewhat flat for most samples (13 out of 15). In some samples the rising tone was artificial and unnatural. In other cases, the intonation peak came too early or falling intonation was too steady and gradual.

Regarding sentence prominence and word stress, the native listeners commented that many L2 speech samples (12 out of 15) showed

inappropriate or indiscernible stress patterns. Also in some cases, too many syllables were stressed, or only primary stress but not second stress was perceptible. Sometimes it was difficult to hear stress although some syllables were longer than others. One of the participants showed a tendency to stress first syllables regardless of various stress patterns of English words. Another problematic pronunciation feature was reduced speech. All of the speech samples showed little to no reduction of unstressed syllables with each syllable carefully and forcefully enunciated. During the production experiment the L2 speakers were asked to read the passage at a natural rate. Although their speech rate was not unnaturally slow, they often overemphasized individual syllables.

On the other hand, the nonnative listeners did not indicate these kinds of problems in the comments section. It seems that the nonnative listeners were less sensitive to a foreign accent involving suprasegmental features than native listeners when assessing L2 speech. For example, Figure 2 shows a spectrogram with the pitch contour of the sentence *Are you a native speaker of English?* by a Korean speaker.

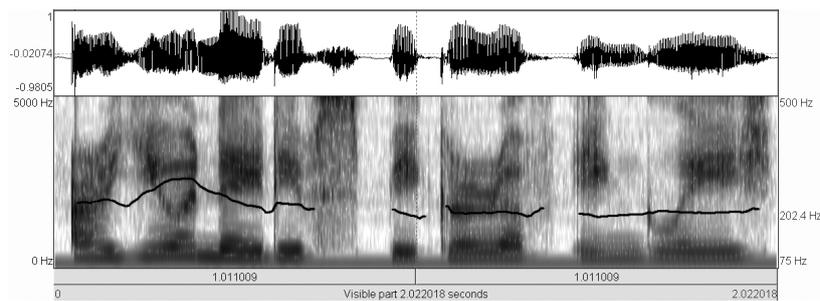


Figure 2. A spectrogram with the pitch contour of the sentence, *Are you a native speaker of English?* (Korean speaker)

Although the sentence is a yes/no question, a rising pitch contour is not shown at the end. The two native listeners' ratings for this sentence were very low (score 2 from both listeners) and they pointed out that the sentence sounded unnaturally flat. However, both a self-rating and a peer-rating for this were high (score 5 and 4, respectively), and the nonnative listeners commented that there were no problems in the intonation pattern of this sentence.

5. Discussion and Conclusion

The present study intends to compare native and nonnative listeners' assessment of L2 English speech. To determine whether there are any mismatches between the assessment of native listeners and that of

nonnative listeners, two groups of listeners (native English listeners and native Korean listeners) rated Korean speakers' speech production in terms of various aspects of pronunciation based on the reading of an English passage.

The results show that although all the Korean speakers who participated in both production and evaluation are advanced learners of English, their rating is different from that of the native English listeners. The differences are clearly shown in their evaluation of suprasegmental features such as intonation, stress and reduced speech. The mean scores given by native listeners turned out to be significantly lower than those of nonnative listeners with regard to these features. This could be interpreted as nonnative listeners having more difficulty perceiving the suprasegmental irregularities of foreign accents than native listeners. If nonnative listeners could not perceive the peculiarities of intonation or stress patterns in L2 speech, it would be more difficult to utilize such properties appropriately in production.

Considering that in L2 speech pronunciation prosody or suprasegmental features play a greater role than segmental sounds, nonnative listeners should pay more attention to suprasegmental aspects. Recent research has revealed that accurate suprasegmental production can increase the intelligibility of nonnative speakers (Anderson-Hsieh et al. 1992, Derwing and Rossiter 2002, Hahn 2004, Boula de Mareuil and Veru-Dimulescu 2006) and that nonnative speakers derive more benefits from instruction in suprasegmental features than from segmental sound instruction (Derwing and Rossiter 2003).

The discrepancies in L2 speech assessment between native listeners and nonnative listeners shown in the present study are not in complete accord with previous studies (Flege 1988, MacKay et al. 2006), where native and nonnative listeners' ratings are highly correlated. Flege (1988) and MacKay et al. (2006) argue that nonnative listeners are able to gauge foreign accents in English much like native listeners. However, the researchers in the previous studies examine the overall degree of foreign accent and they do not consider several pronunciation features separately. The present study compares native listeners' assessments with those of nonnative listeners from various aspects of pronunciation, and a difference between the two groups of listeners is clearly shown in the assessment of suprasegmental features.

The results of the present study also indicate that nonnative listeners' ratings are slightly higher than those of native listeners for all aspects of pronunciation except individual sounds. This difference could be the result of effects stemming from the shared native language background of the participants of production experiment and the nonnative evaluators. Previous studies (van Wijngaarden 2001, Bent and Bradlow 2003, Jongman and Wade 2007) have found that for native listeners, the intelligibility of native speakers is greater than the intelligibility of

nonnative speakers, but that for nonnative listeners, the intelligibility of high proficiency nonnative speakers is equal to, or greater than, the intelligibility of native speakers². In the present study the foreign accent of L2 speech from nonnative English speakers is less salient to nonnative listeners than to native listeners because nonnative speech production and perception are both systematically linked to the same native language sound structure. The discrepancies shown in the assessment between the two listener groups might be more obvious if L2 speakers did not evaluate their own pronunciation. Self-evaluation tends to be stricter than peer-evaluation because many Korean learners of English are sometimes too modest about their English pronunciation. This tendency is apparent in the present study.

It should be noted that as Munro and Derwing (1995) indicate, the relationships between intelligibility and accentedness are partially independent. Although greater accentedness usually reduces intelligibility, the situation is not always straightforward. If L2 utterances are unintelligible, the speech samples are perceived as heavily accented. However, some L2 utterances could be strongly accented but still intelligible. Thus, the accentedness of L2 speech cannot be directly interpreted as unintelligibility.

The present results also have some implications for teaching English pronunciation. Nonnative instructors should be aware of how strongly suprasegmental features affect native listeners' assessment of L2 speech because the lack of suprasegmental awareness and usage can be the most detrimental factor to intelligibility. Thus, the focus on suprasegmental features should be increased in L2 pronunciation instruction. Since previous studies show that suprasegmental aspects of pronunciation can be improved through instruction (Derwing et al. 1998, Pennington and Ellis 2000), more practical methods for improving these aspects should be devised and incorporated into L2 instruction. Moreover, nonnative instructors could benefit by being trained to utilize elements of suprasegmental features of pronunciation even if they already display high proficiency in English.

There are some limitations in the present study. The number of listeners assessing the L2 speech is small and therefore, the findings should not be overgeneralized. Furthermore, the reading task suffers from a degree of artificiality and may not provide the most natural evidence of a speaker's pronunciation. Also, regardless of instructions given, reading tasks necessarily involve careful pronunciation, which cancels out the effect of

² One reviewer pointed out that in a previous study (Bent and Bradlow 2003) interlanguage intelligibility between speakers and listeners was examined based on discourse, and that the results of the study could not be used to interpret those of the present study. However, Bent and Bradlow employed the Northwestern University Foreign Accented English Speech Database containing recordings of 2048 syntactically simple sentences read by 32 non-native speakers. Thus, it can be said that in both their study and the present study similar types of materials were used.

some of the pronunciation features examined (e.g., reduction) to some degree. It may be valuable to obtain a spontaneous sample of spoken speech from L2 learners in future research.

Appendix A. Diagnostic passage³

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. 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 form⁴

Intonation

Listen for any **yes/no questions** in the recording. First, write down two examples and mark her intonation. Second, rate the sound of her accent on a scale of 1-5. Finally, make some comments about her intonation: was it appropriate? Why/why not?

- a. _____
b. _____

³ This passage is a modified version of the one found in a course book (Celce-Murcia et al. 2010: 481).

⁴ This is a slightly modified version of one found in a course book (Celce-Murcia et al. 2010: 487).

- JENKINS, JENNIFER. 2002. A sociolinguistically based, empirically researched pronunciation syllabus for English as an international language. *Applied Linguistics* 23(1), 83-103.
- 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.
- MACKAY, IAN. R. A., JAMES E. FLEGE, and SATOMI IMAI. 2006. Evaluating the effects of chronological age and sentence duration on degree of perceived foreign accent. *Applied Psycholinguistics* 27, 157-183.
- 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.
- PENNINGTON, MARTHA C. and NICK C. ELLIS. 2000. Cantonese speakers' memory for English sentences with prosodic cues. *The Modern Language Journal* 84(3), 372-389.
- THOMPSON, IRENE. 1991. Foreign accents revisited: The English pronunciation of Russian immigrants. *Language Learning* 41(2), 177-204.
- VAN WIJNGAARDEN, SANDER J. 2001. Intelligibility of native and non-native Dutch speech. *Speech Communication* 35, 103-113.

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

received: March 5, 2011

accepted: April 15, 2011