Acoustic Measurement of English read speech
by native and nonnative speakers

Choi, Hansook[1]

ABSTRACT

Foreign accent in second language production depends heavily on the transfer of features from the first language. This study examines acoustic variations in segments and suprasegments by native and nonnative speakers of English, searching for patterns of the transfer and plausible indexes of foreign accent in English. The acoustic variations are analyzed with recorded read speech by 20 native English speakers and 50 Korean learners of English, in terms of vowel formants, vowel duration, and syllabic variation induced by stress. The results show that the acoustic measurements of vowel formants and vowel and syllable durations display difference between native speakers and nonnative speakers. The difference is robust in the production of lax vowels, diphthongs, and stressed syllables, namely the English-specific features. L1 transfer on L2 specification is found both at the segmental levels and at the suprasegmental levels. The transfer levels measured as groups and individuals further show a continuum of divergence from the native-like target. Overall, the eldest group, students who are in the graduate schools, shows more native-like patterns, suggesting weaker foreign accent in English, whereas the high school students tend to involve larger deviation from the native speakers' patterns. Individual results show interdependence between segmental transfer and prosodic transfer, and correlation with self-reported proficiency levels. Additionally, experience factors in English such as length of English study and length of residence in English speaking countries are further discussed as factors to explain the acoustic variation.

1. INTRODUCTION

One of the main goals of foreign language education is mastery of the target language. Both the educator and the learner want to develop better skills in second language (L2) production and perception. Proficiency in a foreign language is assessed in terms of multiple factors such as vocabulary, pronunciation, grammar, fluency, etc. (e.g., Higgs and Clifford, 1982; Iwashita et al, 2008), and phonological skill is discussed generally with degrees of foreign accents or native-like acoustic variations (Baker

et al., 2011; Iwashita et al, 2008 among others). However, the pronunciation accuracy is not found as a direct index of the learner's global proficiency level in the impressionistic study by Iwashita et al. (2008). The current study is designed to compare acoustic variations of native speakers and nonnative speakers with various proficiency levels, and to identify variables that specify different accuracy and fluency levels of nonnative speakers.

Major challenge in learning a second language (L2) is the system that is not identical to the first language (L1). The L1 transfer causes nonnative-like foreign accent in the L2 production (Ioup, 2008; Piske et al., 2001 for a review), and strongly foreign accented speech results in low intelligibility and thus, lower ratings by native listeners (Sebastian et al., 1980; Tajima et al., 1996 out of many). Foreign accent is found to relate to the different phonemic and phonetic patterns at the segmental level (Flege, 1991 & 1995; Flege & Eefting, 1988; MacKay et al.

2001, among others), or the dissimilar prosodic systems at the suprasegmental level (Archibald, 1997; Lee et al., 2007; Zhang et al., 2008) in L1 and L2. The new L2 system is acquired through category assimilation or dissimilation induced by the L1 system according to the Speech Learning Model (Flege, 1995), which accounts for changes across the life span in speech learning. One notion derived from the model, the feature hypothesis, further suggests that L2 features not contrastive in L1 will be more difficult to acquire for the L2 learner and this difficulty will be reflected in the learner's production (McAllister et al, 2002).

English phonology and Korean phonology reveal different features in diverse ways. First, the two languages employ different phonemic features. Most of consonants and vowels are not produced from identical articulation places with same gestural movements. For example, English labio-dental /f/ and /v/, and interdental /θ/ and /ð/ are not employed in Korean consonant system, and English tense and lax contrast in vowels is not specified in Korean vowels. These segmental differences predict L1 transfer and plausible foreign accents in Korean learners' English speech.

Second, Korean employs different phonological rules in addition to the different segments. Korean palatalization before the high front vowel, for example, conditions complementary distribution of /s/ and /ʃ/. English /s/ and /ʃ/, however, mark contrast as in /see/ and /she/. Suppression of Korean palatalization is inevitable to mark the given contrast in English speech without foreign accent.

Third, different syllable structures in Korean and English add further difficulties to the second language learners. For instance, consonant clusters are avoided in Korean, but they are very popular in English. The diphthongs with two transitional vowel qualities are eligible combinations in a single syllable for English whereas the similar sequences are realized as two separate syllables in Korean. Transfer of the Korean syllable structure is very plausible for sequences of multiple consonants and vowels in a single English syllable.

Another difference between two languages is their dissimilar rhythmic resolutions. Although it is not without controversy (Cho, 2004; Han, 1964, Ji, 1993; Lee, 1982; Seong, 1995 among others), it is generally agreed that the two languages are not in the same rhythmic group, and that English is more likely identified as stress-timed whereas Korean is more likely marked as syllable-timed (Abercrombie, 1967 for example). The initial definitions on the typology of speech rhythm idealized that a stress-timed language like English involves an equal distance

between stressed syllables whereas the isochronous units of syllable-timed Korean is syllables (Abercrombie, 1967; Pike, 1945 among others). Though the idea of isochrony is weakened, stress-timed languages are found to specify heavy stressed syllables in comparison to the reduced unstressed syllables (Dasher & Bolinger, 1982; Dauer, 1983 for example). The reduction of the nonprominent syllables is revealed as attenuated gestural movement and duration whereas the prominent syllables gain extra durations and hyper-articulations (de Jong, 1995). Syllable-timed languages, in contrast, involve no severe reduction in nonprominent syllables, keeping equidistance between syllables.

The interference between different L1 and L2 prosody predicts different temporal resolutions in syllables and words by native speakers and L2 learners. The idea of prosodic transfer in L2 studies is relatively new, and there are not so many studies reporting significant evidence of prosodic transfer. For example, the L2 study on the temporal variation by Wang & Behne (2004) reports an insignificant difference in English syllable production between native speakers and Mandarin Chinese learners. However, Lee et al. (2007) and Zhang et al. (2008) reported the influence of native language prosody in English stress marking by Japanese, Korean and Mandarin speakers. The previous research supporting prosodic transfer is still limited in the sense that the variation is discussed for nonnative speakers as one homogeneous group with little attention on individual variations within the nonnative group. Individual variation is discussed in terms of the fluency levels (Cucchiarini et al., 2000; Ramus & Mehler, 1999; Shih & Wu, 2011). However, the findings are on the global facts such that the speech by nonnative speakers who are not fluent is slower at rate with longer vowel duration and more filled pauses (Shih & Wu, 2011). The current study is designed therefore to discuss the language-specific transfer at prosodic levels as well as segmental levels with a focus on different phonological proficiency levels of nonnative speakers. Phonological proficiency in this study includes the accuracy of the segmental production and fluency of the prosodic features.

The first research question of this study is how the proficiency is reflected in vowel specification of Korean learners of English. The feature hypothesis predicts difference between native speakers and nonnative speakers in producing vowel qualities dissimilar in L1 and L2. Acoustic variations of nonnative speakers are compared with those of native speakers as both group and individual. Vowel qualities are examined in terms of formant values. The target tokens are analyzed from continuous speech with consideration of assimilatory effects from neighboring

segments. It is hypothesized that nonnative speakers with lower phonological proficiency levels will involve greater deviation from formant values of native speakers. The deviation will be robust for the English features that are not employed in Korean, suggesting segmental transfer of Korean features.

The second research question is how the phonological proficiency is reflected in the temporal resolutions of the English speech by Korean learners. Prosodic transfer predicts difference in temporal resolutions between native speakers and Korean learners of English. Particularly, Korean learners and native English speakers will demonstrate dissimilar durational patterns for various vowel types and syllables with prominence variation. Phonological proficiency levels are predicted to correspond with the degree of differences such that greater deviation will be observed from speakers with lowest phonological proficiency levels.

The third research question is whether there is a correspondence between phonological proficiency at the segmental level and at the prosodic level. The paradoxical aspect of prosody in L1 acquisition is that language specific-prosody begins to be recognized by very young infants like newborns (e.g., Mehler et al., 1988), but full-acquisition is relatively later in comparison to segmental acquisition (e.g., Cutler & Swinny, 1987). The question is how prosodic features and segmental features interact in L2 acquisition. This study examines whether deviation from the target in formant values shows an agreement with deviation in temporal resolutions, given that the deviation from the native target, the non-nativelikeness, is taken as an index of L2 phonological skills. A strong correlation between segmental and temporal deviations will suggest a holistic and interactive development of segmental and suprasegmental proficiency of the second language.

The final goal is to find plausible speech-external factors that can explain the phonological proficiency in the speech production of Korean learners of English. The phonological proficiency levels measured from acoustic outputs are compared with self-reported English levels and pronunciation accuracy levels. Other background features will be explored in relation to the measured phonological proficiency levels in order to find out the factors to explain the acoustic variation of nonnative speakers. The given research goals will be discussed with the following methods.

2. METHOD

2.1 Speech Material and Subject

The speech data is extracted from Prawn-DB[2]. The speech corpus includes the read speech recordings on the identical

sentences and words by native English speakers and Korean learners of English. The current study analyzes the speech by 20 native English speakers, 20 Korean high school students, 10 Korean undergraduate students, 20 Korean graduate students. More detailed information of individual speakers is additionally provided in the Prawn-DB, which is considered as additional speech-external factors for the acoustic variation in speech production. For the native English speakers, their nationalities, gender, age, length of residence in Korea, location of stay in Korea, and birthplaces can be found. Most of the native speakers were from USA and Canada, and two British English speakers and one Australian speaker were included in the native speaker groups. Information of Korean learners includes speakers' age, gender, hometown, parents' birthplaces, numbers of years learning English and staying in the English-speaking countries, self-judged English proficiency level, self-judged level of English pronunciation accuracy, and interest level in the English language. These self-reported values are analyzed as plausible factors to explain the variation in the acoustic productions.

Speech token are from a read speech corpus by the target subjects. The sentence of "*My daughter will outgrow these shoes soon*", read by all the participants, is used for the analysis of vowel formants and duration. Another sentence of "*Anna is a great doctor in Canada, isn't she? Well, Anna still is a great doctor*" is used for durational measurement at the syllable level. The syllable variation is discussed with the three syllables in '*Canada.*' Some speakers read a variation of "*Anna is a great doctor in Canada, isn't she? Well, Anna is a great doctor still now.*" The variation was read by three native speakers and all of the high school students. Additionally, a phrase, "pop up", is used to measure a lowest point vowel for formant transformation. The tokens for the analysis are selected from rather lengthy sentences. The target sentences were produced with more natural intonation with greater emphatic variations.

2.2 Measurements

2.2.1 Formant Measurement

Vowel quality is compared in terms of the first formant (F1) and the second formant (F2) in this study. The linear formant values are measured in the corresponding vocalic portion using Burg method in Praat at the time step of 0.001 second, with maximum formant at 5500 Hz[3], and with the analysis window of

---

2) Chung et al. (2008) provides more detailed information of the data base.
3) For male speakers, additional measurement was performed with

0.025 second (Boersma, P. & Weenink, D., 2010). The vocalic portion of the vowel target is hand-labeled based on the wave form and formant transitions. The onset of glottal pulsing with formant patterns is marked as the vocalic onset and cessation of pulsing as the end point, and the transitional points are labeled when the target is between sonorants. The F1 and F2 are measured at the mid-point of the vowel interval. 5 vowels (/ɔ/ in daughter, /ɪ/ in still, /o/ in grow, /i/ in these, /u/ in soon) were measured for F1 and F2 and transferred to the normalized values following the S-centroid method.

For the point vowel measures in S-centroid method, the lowest vowel, /ɑ/, is additionally analyzed from a read speech token of "pop up", at the mid point of the vocalic interval. Though British speakers and American English speakers may use different vowel qualities for the target, it is taken to get the lowest vowel formant values regardless of the backness of the vowel.

2.2.2 Formant normalization

Formant values depend on variable factors. For example, formant values are gender-dependent. Females involve higher F1 and F2 values and young children have higher formant values (e.g., Diehl et al., 1996; Peterson & Barney, 1952; Yang, 1996). The physiological difference in the vocal tract induces the variation in acoustic outputs such as formant values, and several techniques has been suggested to neutralize speaker-dependent features in formant values (Adank, 2003; Labov et al., 2006; Thomas & Kendall, 2007; Zwicker & Feldtkeller 1967). This study takes the S-centroid method to normalize the gender or age variation of speakers (Watt & Fabricius, 2002).

The S-centroid method has recently been employed in studies of variation in a variety of the English language (Bigham, 2008; Fabricius, 2007; Kamata, 2008; Watt & Fabricius, 2002). The S-procedure normalizes a speaker's set of vowel data by expressing each formant value as a proportion of its respective centroid value, which is derived using F1 and F2 maxima and minima for that individual's vowel space. A centroid point, S, is derived from three "point vowels", namely /i/, the most front, /ɑ/, the lowest, and an idealized backmost point, /u'/. Note that the /u'/ vowel is not an observed, but a derived one, defined as F2 (u') = F1 (u') = F1 (i). The idealized backmost point has the same F1 and F2 values since F2 cannot have a lower frequency than F1 by definition, "but often has a frequency so close to it that the spectral peaks cannot reliably be distinguished from one

another using instrumental analysis, we can justifiably assume for present purposes that the speaker's closest, backest possible vowel has an F2 exactly equivalent to its F1 frequency." (Watt & Fabricius, 2002: 162). The formula for S is as follows:

$$S(Fn) = ([i]Fn + [ɑ]Fn + [u']Fn)/3$$

All the observed measurements of Fn are then divided by the S value for that formant n, and all resulting figures are expressed on the scale of Fn/S(Fn).

2.2.3 Measurement of Duration

As separate acoustic cues for the English proficiency, durational values were measures with respect to three different levels, vowel, syllable, and word. The target vowels were all stressed, and included monophthongs and diphthongs together. A diphthong vowel is included to diversify the durational patterns in vowels. The novice Korean learners of English have good chance of producing the diphthong with additional duration by identifying it as a disyllabic target.

Vowel duration is measured from the sentence of *"My daughter will outgrow these shoes soon"*, and the vocalic intervals are measured corresponding to /aj/ in 'my', /ɔ/ in 'daughter', /i/ in 'these', and /u/ in 'soon'.

Syllable duration is measured to see the dissimilar temporal resolution between stressed syllables and unstressed syllables. The stressed target is measured in the word, 'Canada.' The durations of the initial syllable, /kæ/ is compared with the second unstressed syllable, /nə/, and the total duration of the word with three syllables, /kænədə/. The duration of /kæ/ is measured from the cessation of previous periodic period before the stop release to the offset of the voicing preceding the nasal formants. The duration of /nə/ is measured from the onset of the nasal formant to the end of the clear voicing pulse before the weakened closure pulse for the voiced alveolar stop. In the same way, the end of clear pulsing with continuation of previous formants of /də/ is marked for the end of the word, and the beginning of /kæ/ is taken as the beginning of the word. The duration between the beginning and the end of the word is measured as word duration.

Based on the raw durational values, the relative duration of the prominent first syllable is calculated into two ratios. One is the ratio of /kæ/ vs. /nə/, which is calculated by dividing the duration of /kæ/ by the duration of /nə/. The other is the ratio of /kæ/ vs. /kænədə/, calculated by dividing the duration of /kæ/ by the duration of /kænədə/. Greater values in ratios indicate greater

---

maximum formant at 5000Hz. The results have no big difference at low frequencies with the results reported here.

durational strengthening of the prominent syllable overall.

2.3 Statistical Analysis

The deviation of diverse speaker groups is analyzed using Analyses of Variance (ANOVAs). Speaker groups are divided into 2 native speaker groups and 3 nonnative speaker groups. The speaker groups are divided based on similarities in patterns of the target language acquisition, though the graduate students have wide variation of ages. The native groups consist of one for speakers from USA and Canada and the other for British and Australian speakers. The tokens from continuous speech show effects from the contexts, and thus, all individual formant values are treated separately in the analysis. Pearson correlation coefficient measurement is employed to observe interrelations between diverse factors. For more detailed analysis of the observed factors, post hoc comparisons were performed at the significance level of 0.05. The PASW statistical package (PASW Statistics 18, Release 18.0.0, 15 Jul 30 2009, Polar Engineering and Consulting) was used for the statistical measurement.

In addition to statistical comparisons, distributions of individual subjects' tokens in each condition were compared using diverse charts and graphs.

3. RESULTS AND DISCUSSION

3.1 Vowel Formant

Statistical analyses revealed that most of the normalized formant values are significantly different across the speaker groups. One way ANOVA for the factor of speaker group is employed to observe group distinction, where the speakers are grouped into 5 types of speakers: 2 native speakers (North American vs. non-North American), nonnative high school students, nonnative undergraduates, nonnative graduates. For the significant differences, a post hoc test using Tukey HSD multiple comparisons at (alpha=0.05) are conducted. The ANOVA results are provided in <Table 1>. Significant differences are found consistently in the lax English vowels. Both F1 and F2 of /ɪ/ and /ɔ/ report significant variation due to speaker groups. The group distinction in manifesting the tense vowels is relatively less robust.

More detailed comparison reveals the foreign accent shared by Korean speakers in their English vowel production. Tukey HSD multiple comparison (alpha=0.05) shows that the F1 values of /ɪ/ in 'still' by the native speaker groups are significantly different from those by all nonnative groups. The native speakers produce significantly higher F1 suggesting more opening of mouth. In F2 of /ɪ/, the group distinction is mainly from the North American native speakers who produce significantly lower F2 values than all the nonnative speakers. The pattern suggests that /ɪ/ in 'still' is more centralized in native speakers' speech in comparison to the one by nonnative speakers. The centralization of native speaker groups is clear in <Figure 1(a)>. Note that the F1 and F2 scales are the converted values using the S-centroid method, and that the scales do not correspond to the physical structure of the oral cavity.

Table1. Statistical results by one-way ANOVA on F1 and F2 of English vowels (*** p<.0001 ** p<.01, * p<.05)

	/ɪ/	/i/	/ɔ/	/o/	/u/
F1	F(4, 65) =12.626 ***	F(4, 65) =4.117 **	F(4, 65) =12.253 ***	F(4, 65) =1.491	F(4, 65) =.486
F2	F(4, 65) =27.813 ***	F(4, 65) =1.466	F(4, 65) =11.026 ***	F(4, 65) =17.725 ***	F(4, 65) =3.430 *

In contrast to the robust deviation in the lax vowel, the tense vowel, /i/, shows relatively minor distinction between native and nonnative groups. The group distinction is not significant in F2, and only F1 shows a significant variation. Post hoc multiple comparison tests reveal that F1 values of /i/ by North American native speakers are significantly lower than the values of high school students (at alpha=0.05). As shown in <Figure 1(b)>, the native speakers produce tokens with even smaller F1 values, which is in the Y-axis, suggesting higher tongue position. The separation is observed between square symbols and the flower symbols, denoting high school students and North American native speakers respectively. However, the separation between native and nonnative speakers in <Figure 1(b)> is not as obvious as in the separation in <Figure 1(a)>.

The location of target vowels in <Figure 1(a) & (b)> visualizes difference between tense and lax high front vowels in the acoustic space, which is obvious for the native speakers. The targets of nonnative groups are not very distinguishable between the two different vowel qualities. The similarity in specifying tense and lax high front vowels suggests the phonemic transfer of Korean vowel system. Korean speakers tend to neutralize the tense and lax contrast in English, and produce a vowel of intermediate values between the target tense and lax vowels, which confirms the results in Flege et al. (1997). The neutralization is more noticeable with high school student groups,

the group of lowest phonological proficiency, whereas graduate students tend to show greater overlap with native speakers.
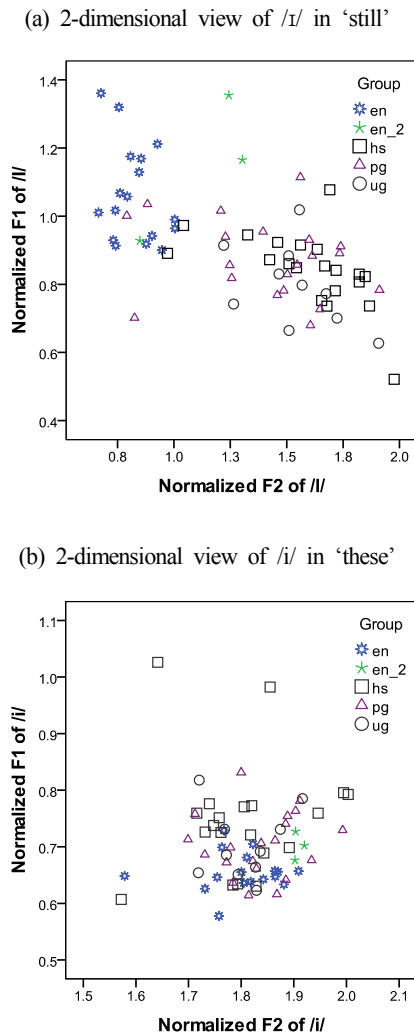
(a) 2-dimensional view of /ɪ/ in 'still'

(b) 2-dimensional view of /i/ in 'these'

Figure 1. Scatter plot of the English front vowels in terms of normalized F1 and F2[4]).

The lax vowel in 'daughter' reveals a significant difference between native speaker and nonnative speakers. Multiple comparison using Tukey Posthoc shows that a significant group distinction exists between North American native speakers and all nonnative in their F1 values of /ɔ/. The non-North American native speakers also show significant difference from nonnative high school students and undergraduates. The native speakers

have higher F1 values than nonnative speakers in general suggesting further opening of the lax vowel. The significant group distinction in F2 values of /ɔ/ is mainly due to the marked F2 patterns of non-North American native speakers. Post hoc comparison shows a significant difference between non-North American native speakers and all others (American native, nonnative speakers). The normalized values of non-North American native speakers are significantly higher than the other groups as is visualized in <Figure 2>.

(a) 2-dimensional view of /ɔ/ in 'daughter'

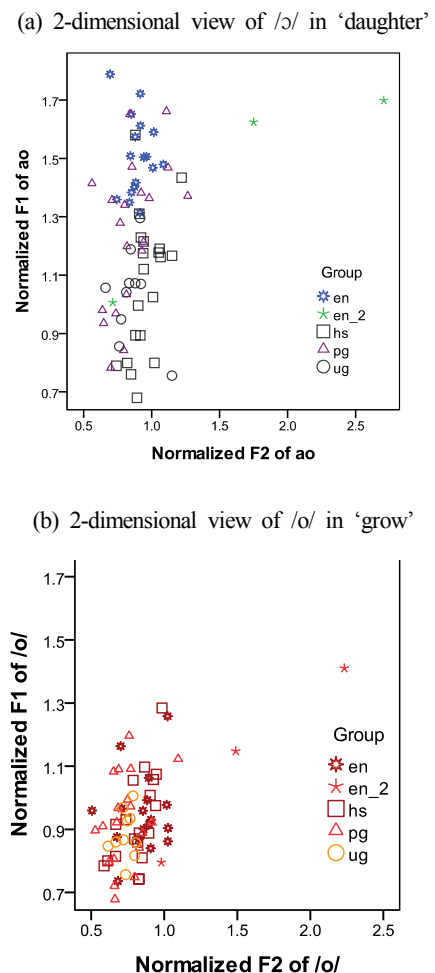(b) 2-dimensional view of /o/ in 'grow'

Figure 2. Scatter plot of the mid back tense and lax vowels in terms of normalized F1 and F2.

In fact, the two British speakers produced significant different vowels. The two outliers, marked with star symbols in <Figure 2> are the tokens by the British speakers in the data. The other native speakers still display obvious difference from the nonnative speakers in <Figure 2>, in that there is a clear distinction between the two vowels especially in the F1 dimension. The lax one in <Figure 2(a) > involves a lot higher f1 values than the

---

4) In <Figure 1 & 2>, symbols marked 'en' are for North American native speakers, 'en_2' for the British and Australian native speakers, 'hs' for high school students, 'pg' for graduate students, and 'ug' for undergraduates. The scale bases on the normalized F1 and F2. Higher F1 is for greater opening and higher F2 is for further fronting, which is in the up and right direction here.

tense one in <Figure 2(b) >. The nonnative speakers, on the other hand, specify the tense and lax vowels in the similar acoustic dimension. The Y-axis for F1 is scaled identically in <Figure 2(a) and (b)>. The tokens for native speakers display a clear transition when the vowel quality is changed especially in the Y-axis. The difference is not very noticeable for the nonnative speakers' tokens.

Another significant distinction is also due to the distinctive patterns by the British speakers. F2 of /o/ and F2 of /u/ show significantly higher values by the native speakers from non-American countries. The North American native speakers and nonnative speakers do not reveal a robust difference.
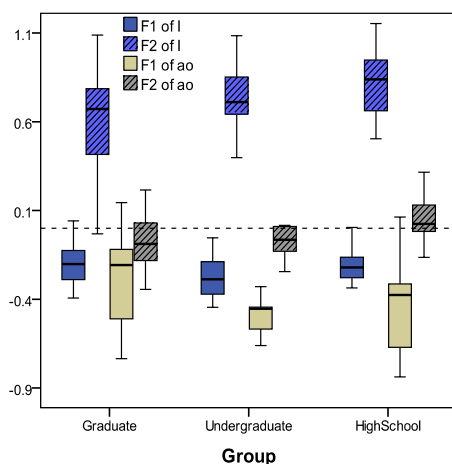

Figure 3. Box plot of the difference in formant values of lax vowels from the mean values of native speakers.

The formant values show distinction between native speakers and nonnative speakers particularly in producing lax vowel targets. More distinction is between the high school students and native speakers, whereas graduate students show greater similarity with the native speaker groups. As an index of phonological proficiency measure, the measured formant values are transferred into difference values by deducting the mean values of native speakers from the individual native speakers.

Mean values of native speakers are calculated only from the North American speakers in order to minimize outliers and enhance the homogeneousness of the native speaker groups. The other native speakers are not included. The difference values further reveal the distinctions of nonnative speaker groups.

Significant difference among nonnative speaker groups is found in F2 of /ɪ/ (F(2,49)=4.295, p<.05), F1 of /ɔ/ F(2,49)=3.964, p<.05) and F(2,49)=3.177, p=.051). The graduate students show the smallest deviation from the native speakers, whereas the high

school students show the greatest deviation, which is depicted in <Figure 3>. The dashed line in the middle is the mean of the native speakers. From the dashed line, graduates take less distance compared to the other groups. High school students tend to depict greater distance from the middle. Given that the distance is taken as phonological proficiency index, graduate students can be considered as a group with the highest phonological proficiency.

3.2 Temporal resolution of syllables and vowels

One-way ANOVA test on the mean vowel duration reports that the group distinctions are not significant (F(4, 65)= 1.290, p=0.283). In contrast to the previous studies that suggest the longer vowel duration for nonnative speakers' speech, the current study does not find any clear distinction between groups in the mean vowel duration of the four vowels. The group is still a significant factor to differentiate the standard deviation (SD) of vowel durations (F(4, 65)=4.756, p<0.01). Tukey posthoc comparison at 0.05 level reports that SD was significantly different between North American native speakers and high school students, and between graduates and high school students. High school students have significantly greater SD of vowel durations. As is found in the formant measurements, the high school students are found as the group of the lowest phonological proficiency in temporal resolution of vowels. Among the target vowels, the duration of the diphthong, /aj/ display the significant group distinctions among nonnative speakers (F(4, 65)=7.853, p<.001). Posthoc comparison reveals the significantly longer duration of the high school students (alpha=0.05). The native groups produced the smallest duration for the diphthong. The results suggest that the nonnative speakers tend to produce the diphthong with extra duration and the lengthening is most obvious in the high school student group.

Syllable variation due to dissimilar prosodic prominence contexts also shows the difference between native speakers and nonnative speakers. Different proficiency levels in nonnative groups are also suggested from the multiple comparisons between groups. One-way ANOVA on the ratio of the duration of 'ca' vs. the duration of 'na' in 'Canada' shows a significant group distinctions (F(4, 65)= 9.525, p<0.001). The ratio of 'ca' vs. the word duration of 'Canada' also significantly different in the given groups (F(4, 65)=13.594, p<0.001). Tukey posthoc comparison at 0.05 level reports that the native speakers particularly from North America produce higher ratios than nonnative groups do in both measurements. Among the nonnative groups, high school students

always mark the lowest ratios. The native speakers produce the stressed syllable with extra duration compared to the following unstressed syllables, while the nonnative groups tend to produce less variant syllable durations. The stress-induced durational lengthening that is more robust in native speakers confirms the prosodic transfer in Korean learners of English. Korean speakers from a less stress-timed language tend to minimize the variation of syllable duration. The prosodic transfer also reflects dissimilar phonological proficiency levels. The lengthening of the stressed syllable is smaller in the high school students' tokens and larger in the graduate students. The phonological proficiency of groups suggested from the formant patterns is further confirmed by temporal resolutions.

Scatter plots in <Figure 4> depicts the separation of native speaker groups and nonnative speaker groups in specifying vowel durations and stressed syllable durations. The right bottom portion of the acoustic space, with higher ratio and lower SD and diphthong duration values, is crowed with the native speakers' tokens. The tokens for the high school students are in general farther from the space. The individuals, who are away from the native speakers' tokens in the Y-axis, the ratio values, are placed away from native-like vowel duration manifestation.
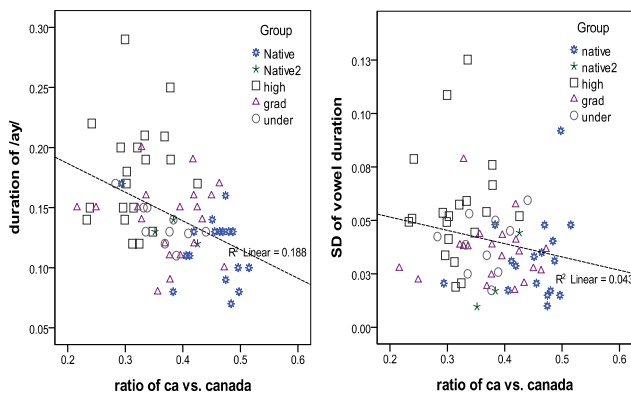

Figure 4. Scatter plots of durational measurements.

<Figure 4> also reveals the reliability of the difference durational measurements as indices of individual phonological proficiency levels. A good correlation exists between the ratio values and the vowel durational values such that the speakers who are making higher ratio produce less variation in different vowel structures and shorter duration for the diphthong. In turn, the speakers with less stress-induced lengthening produce the diphthong with extra duration and more durational variation due to different vowel structures. The transfer of the Korean prosody involves all the measured durational features of English

production in a similar way. The results suggest that different levels of prosodic proficiency of L2 seem to develop interdependently.

3.3 Interdependency of phonological proficiency indices

The difference between native and nonnative speakers is revealed in the segmental specification of vowel equality and prosodic specifications of vowel duration and stressed syllable duration. The predicted L1 interference is detected in L2 production in the current study. The different phonological proficiency levels of Korean learners are also identified from the different acoustic measurements. However, the phonological proficiency is discussed as groups of different educational levels and ages in the previous sections. It still needs to be discussed whether the individual phonological proficiency levels show similar patterns across the diverse acoustic measurements. In order to measure the individual phonological proficiency levels, the deviation from the native patterns is measured for each measurement of nonnative speakers. The deviation values were the deduction outputs from the means of stereotypical native speakers, namely the North American speakers without outliers. Each mean native value was deducted from corresponding nonnative values of individual Korean speakers and the results were taken as indices of acoustically measured phonological proficiency levels. A Pearson correlation coefficient analyses were employed to observe interdependencies among the diverse phonological proficiency values. Results are summarized in <Table 2>.

The correlation coefficient values in <Table 2>suggest that the phonological proficiency measures from diverse acoustic cues depend on one another for each individual speaker. Individual speakers' deviation values show significant correlations between the durational values and segmental values of formant deviations. The correlation suggests that the different acoustic measurements for speech-internal phonological proficiency identify the individual learner's level in a similar way. For example, the negative correlation between 'Ca vs. na' and F2 value in some lax vowels implies that the Korean learners who are making smaller exaggeration of the stressed syllable are marking bigger F2 values or more peripheral vowel qualities than the native targets. So, the lax vowels tend to produce more similar to tense vowels in such speakers particularly in protrusion. The positive correlation between 'ca vs. na' and the F2 of the tense vowel indicates the opposite pattern. The Korean learners who do not mark stress-induced exaggeration tend to produce less peripheral

vowels, not reaching the target front vowel locus. It is, however, true that the correlation is not one by one, namely between every single formant feature and every single durational feature. Certain segmental features involve significant interaction with some durational features. The results can be due to the limited size of sound tokens observed in the current study. Otherwise, it suggests that assessment of phonological proficiency should base on multiple acoustic cues for reliable results.

Table 2. Pearson correlation coefficient between vowel qualities and durational measurement (** p<.01, * p<.05, 2-tailed)

	Ca vs. na	Ca vs. word	SD. Of Vdur	Dur of /aj/	Mean Vdur
F1 of /ɪ/	.233	.158	.105	.112	.122
F2 of /ɪ/	-.285*	-.220	.073	.267	-.029
F1 of /o/	.023	.005	-.067	.007	.255
F2 of /o/	-.310*	-.348*	.033	.105	-.026
F1 of /i/	-.091	-.127	.117	.140	-.092
F2 of /i/	.400**	.356*	.172	.327*	.313*
F1 of /ɔ/	.199	.090	-.017	.069	.374**
F2 of /ɔ/	-.198	-.226	.374**	.321*	-.169

The phonological proficiency variation in the speech-internal values also involves an interaction with speech-external factors. The self-evaluation measurements reported by participants show correlation with the acoustic deviation values measured from the speech data. All the nonnative subjects judged their own English proficiency into 5 levels - basic, low-intermediate, intermediate, up-intermediate, and advanced. They also evaluated their interest in English and accuracy in English pronunciation into 5 levels – very low, low, mid, high, and very high. The relationship between the speech internal proficiency levels and the self-reported levels are explored through the Pearson correlation coefficient measurements. Additional information is also included in the analysis in order to find out the explanatory factors for the proficiency levels. The number of years learning English, and the number of years spent in English-speaking countries in addition to the self-judged proficiency levels are transferred into numerical values by putting higher numbers for higher levels. <Table 3> provides Pearson correlation coefficients between proficiencies of acoustic features and other reported variables.

In <Table 3>, vowel quality difference for diverse formant differences is combined as one value by adding all squared values of difference measurements of each formant value. The negative correlation values between the speech-external, self-reported values

and speech-internal values of formant differences, /aj/ duration, and SD of vowel duration indicate that the higher levels or longer years interact with less formant difference, smaller SD Of vowel duration, and shorter diphthong specification. On the other hand, the positive r values with stressed syllable ratio indicate the higher levels with greater enhancement of the stressed syllable duration. The overall direction indicates a good correspondence between speech internal acoustic measurements and speech external self-reported variables, though significant linear correlations are found from only some of the interactions. Although the self-reported English levels do not show any significant linear correlation with the acoustic measurements, the bar graph using mean values for each English level in <Figure 5> still shows a tendency of higher proficiency of acoustic measurements in accordance with the self-reported English proficiency levels.

Table 3. Pearson correlation coefficient between acoustic measurements and individual indexes (** p<.01, * p<.05, 2-tailed)

	Formant Difference	Duration of /aj/	SD Of Vowel duration	Stressed syllable ratio
Year Study	-.446**	-.216	-.162	.192
Year in E country	-.207	-.067	-.157	.261
English level	-.216	-.233	-.216	.156
Interest English	.209	-.390**	-.174	-.074
Pronunciation Accuracy	-.279*	-.293*	.018	.039

Based on the significance of correlation coefficient, some speech-external factors can be suggested more significant in explaining the foreign accents in English production by Korean learners. In general, the number of years spent in English-speaking countries does not seem a direct factor for the proficiency in the current acoustic measurements. Self-reported English proficiency levels are not a faithful indicator for the phonological proficiency measured in the speech data either. Stronger influences are from the number of years studying English, self-reported pronunciation accuracy levels, and the level of interest in English, on the other hand. It is not very surprising that the learners who have studied English longer demonstrate more native-like patterns. The results may imply that studying English in any countries is effective in minimizing foreign accent. Note that the number of years learning English does not directly

reflect the subjects' educational status. Majority of the high school subjects responded as 9 years whereas the majority of undergraduate responded as 8 years. Self-reported pronunciation accuracy levels and the levels of interest in English still correlate with vowel formant patterns and diphthong duration. The pronunciation accuracy shows a strong correlation with the vowel formant measurements, and thus Korean learners seem to evaluate the accuracy of their English pronunciation more from the quality of sounds they produce rather than the quantity or durational patterns of sounds.
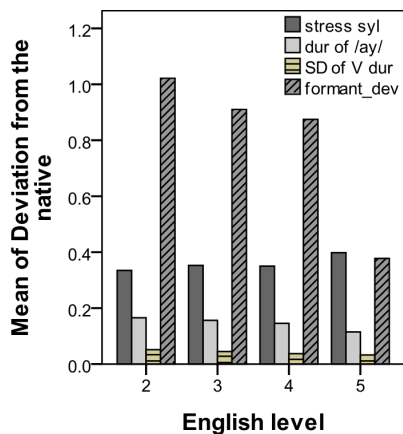

Figure 5. Acoustic measurements in terms of self-reported English proficiency level.

To sum, the acoustic measurements of vowel formants and vocalic and syllabic durations reveal the difference between native speakers and nonnative speakers. The difference is particularly robust in the production of lax vowels, diphthongs, and stressed syllables. The tense/lax contrast is not specified distinctively in English speech by Korean learners. The Korean speakers tend to produce English diphthong with extra duration, and thus, the resulted standard deviation of different vowels is greater than native speakers. The stress-induced lengthening is robust for the native English speaker, whereas the durational variation due to stress is minor for the Korean speakers. The transfer of L1 features is found both at the segmental specifications and at the suprasegmental specifications.

The phonological proficiency levels measured as a group or individuals are realized into a continuum of divergence from the targets by native speakers. The eldest groups, students who are in the graduate schools, generally display most native-like patterns both in vowel formants and vowel and syllable durations, suggesting overall better proficiency in English. On the other hand, the high school students tend to involve larger deviation

from the native speakers' patterns. They are counted as a group of the lowest phonological proficiency. The distinctions between groups show a big picture of rather coarse phonological proficiency levels that can be measured from acoustic features of spoken English. The study also tried to give detailed evaluation points to individual speakers across groups. The acoustic measurements of both segmental features and suprasegmental features show a good agreement in proficiency reports of different groups and individuals.

The proficiency levels measured from the acoustic features also show correlation with the reported variables though the correlation is not always significantly linear. The acoustic patterns of Korean learners of English can be explained partially by the self-reported English proficiency levels and the number of years learning English. The linear correlation does not reveal a strong effect from the experience of staying in English speaking countries.

4. CONCLUSION

The current study provides a supporting example of L1 transfer on L2 specification both at the phonemic levels and at the prosodic levels. The results demonstrate that L1 transfer patterns at different levels have a correspondence among them, which indicate an interactive and holistic development of segmental and suprasegmental proficiency of the second language. This study reveals the heterogeneous patterns of L1 transfer in various Korean learners and suggests the patterns as the L2 proficiency indexes.

All in all the transfer predicted from dissimilar vowel systems, syllable structures, and stress system in English and Korean is detected in the speech by the nonnative speakers. The transfer is dissimilar among different Korean learner groups such that segmental and suprasegmental deviation from the standard is even more obvious in the high school students. The current study, thus, suggests different degrees of deviation from the native speakers' patterns in the identical contexts can be counted as an index of proficiency levels of L2 speakers. The Korean accent in the spoken English is measured by the vowel quality, vowel duration, and stressed syllable duration in this study. The acoustic measurements also involve correspondence with self-evaluations from different sources and other extra-linguistic factors.

The proficiency measurement in this study is based on the difference between English and Korean. Therefore, the future studies on proficiency measurements of different languages should

employ adjusted targets considering the given linguistic contexts. The comparison is also advised to be done with more caution. Segmental realization show a good amount of coarticulation with neighboring segments, and the coarticulatory patterns involve variations due to prosodic contexts. A direct comparison of segmental qualities, simply based on phonemic identification, can mislead the deviational patterns consequently.

REFERENCES

Adank, P. (2003). "Vowel normalization: A perceptual-acoustic study of Dutch vowels", Ph.D. thesis, Katholieke Universiteit Nijmegen.

Abercrombie, D. (1967). *Elements of general phonetics*, Chicago: Aldine.

Baker, R., Baese-Berk, M., Bonnasse-Gahot, L., Kim, M., Van Engen, K., & Bradlow, A. (2011). "Word durations in non-native English", *Journal of Phonetics*, 39, pp. 1-17.

Rachel B., Baese-Berk, M., Bonnasse-Gahot L., Kim, M., Kristin J., Engen, V., & Bradlow, A. (2011). "Word durations in non-native English", *Journal of Phonetics*, 39, pp. 1-17.

Bigham, D. (2008). "Dialect contact and accommodation among emerging adults in a university setting", Ph.D. thesis, The University of Texas at Austin.

Boersma, P. & Weenink, D. (2010). The Praat program, Version 5.1.23, University of Amsterdam.

Cho, M. H. (2004). "Rhythm typology of Korean speech", Cognitive Process, 5, pp. 249-253.

Chung, H., Jang, T. Y., Yun, W., Yun, I., & Sa, J. (2008). "A study on automatic measurement of pronunciation accuracy of English speech produced by Korean learners of English", *Language and Linguistics*, 42, pp. 165-196.

Cucchiarini, C., Strik, H., & Boves, L. (2000). "Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology", *Journal of Acoustic Society of America*, 107, pp. 989-999.

Cutler, A. & Swinney, D. (1987). "Prosody and the development of comprehension", *Journal of Child Language*, 14, pp. 145-167.

Dasher, R., & Bolinger, D. (1982). "On pre-accentual lengthening", *Journal of the International Phonetic Association*, 12, pp. 58-69.

Dauer, R. (1983). "Stress-timing and syllable-timing reanalyzed", *Journal of Phonetics*, 11, pp. 51-62.

De Jong, K. (1995). "The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation", *Journal of Acoustic Society of America*, 97, pp. 491-504.

Diehl, R., Lindblom, B., Hoemeke, K., & Fahey R. (1996), "On explaining certain male-female differences in the phonetic realization of vowel categories", *Journal of Phonetics*, Vol. 24, pp. 187-208.

Fabricius, A. (2007). "Vowel formants and angle measurements in diachronic sociophonetic studies: FOOT fronting in RP", *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, pp. 1477-1480.

Flege, J. (1991). "Age of learning affects the authenticity of VOT in stop consonants produced in a second language", *Journal of the Acoustical Society of America*, 89, pp. 395-411.

Flege, J. (1995). "Second-language speech learning: theory, findings, and problems", In: Strange, W. (ed.) Speech Perception and Linguistic Experience. Timonium, MD; York Press, pp. 233-277.

Flege, J., & Eefting, W. (1988). "Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation", *Journal of the Acoustical Society of America*, 83(2), pp. 729-740.

Han, M. S. (1964). "Duration of Korean vowels", *Studies in the Phonology of Asian Languages II*, University of Southern California.

Ioup, G. (2008). "Exploring the role of age in the acquisition of a second language phonology", In Edwards & Zampini (Eds), *Phonology and second language acquisition*, pp. 41-62, Amsterdam: John Benjamins.

Ji, M. J. (1993). "The duration of sounds" (In Korean), *New Korean Language*, 3, pp. 39-57.

Kamata, Miho. (2008). "An acoustic sociophonetic study of three London vowels", Ph.D. thesis, University of Leeds.

Labov, W., Ash, S., & Boberg, C. (2006). *The atlas of North American English: Phonology, phonetics, and sound change*, Berlin: Mouton de Gruyter.

Lee, B., Guion, S., & Harada, T. (2006). "Acoustic analysis of the production of unstressed English vowels by early and late Korean and Japanese bilinguals", *SSLA*, 28, pp. 487-513.

Lee, H. B. (1982). "A phonetic study on Korean rhythm" (in Korean), *Malsori (Speech Sounds)*, 4, pp. 31-48.

MacKay, I., Meador, D. & Flege, J. (2001). "The identification of English consonants by native speakers of Italian", *Phonetica*, 58, pp. 103-125.

McAllister, R., Flege, J., & Piske, T. (2002). "The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English, and Estonian, *Journal of Phonetics*, 30, pp. 229-258.

Mehler, J., Jusczyk, P.W., Lambertz, G., Halsted, G., Bertoncini, J., & Amiel-Tison, C. (1988). "A precursor of language acquisition in young infants", *Cognition*, 29, pp. 143-178.

Peterson, G. E. & Barney, H.L. (1952). "Control methods used in the study of the vowels", *Journal of the Acoustical Society of America*, 24, pp. 175-184.

Pike, K. L. (1945). *The Intonation of American English*, Ann Arbor: University of Michigan Press.

Piske, T., MacKay, I., & Flege, J. (2001). "Factors affecting degree of foreign accent in an L2: a review", *Journal of Phonetics*, 29, pp. 191-215.

Ramus, F. & Mehler, J. (1999). "Language identification with suprasegmental cues: a study based on speech resynthesis", *Journal of Acoustic Society of America*, 105, pp. 512-521.

Sebastian, R., Ryan, E., Keogh, T. and Schmidt, A. (1980). "The effects of negative affect arousal on reactions to speakers", In H. Giles, W. Robinson, and P. Smith (Eds), *Language: Social Psychological Perspectives,* Oxford: Pergamon Press, pp. 195-213.

Seong, C.J. (1995). "The Experimental Phonetic Study of the Standard Current Korean Speech Rhythm: with respect to its temporal structure (In Korean)". PhD Thesis. Seoul National University.

Shih, C. & Wu, C. H. (2011). "Evaluating Second Language Fluency", *Proceedings of VLSP 2011*, pp. 38-41.

Tajima, K., Port, R., & Dalby, J. (1997). "Effects of temporal correction on intelligibility of foreign-accented English", *Journal of Phonetics*, 25, pp. 1-25.

Wang, Y. & Behne, D. (2004). "Temporal patterns of native Mandarin Chinese speakers' productions of English stop-vowel syllable", *Journal of Acoustic Society of America*, 115, pp. 2393-2393.

Watt, D., & Fabricius, A. (2002). "Evaluation of a technique for improving the mapping of multiple speakers vowel spaces in the F1-F2 plane", *Leeds Working Papers in Linguistics and Phonetics* 9, pp. 59-173.

Yang, B. (1996). "A comparative study of American English and Korean vowels produced males and female speakers", *Journal of Phonetics*, 24, pp. 245-261.

Zhang, Y., Nissen, S., & Francis A. (2008). "Acoustic characteristics of English lexical stress produced by native Mandarin speakers", *Journal of Acoustic Society of America*, 123, pp. 4498-4513.

Zwicker, E. & Feldtkeller, R. (1967). *Das Ohr als Nachrichtenempfänger*, Stuttgart: S. Hirtzel Verlag.

● **Choi, Hansook**
  Research Affiliate (2009~present)
  Haskins Laboratories
  New Haven, CT USA
  Email: choi@haskins.yale.edu