ISSN 2005-8063
2016. 3.31.
Vol.8 No.1
pp. 19-24

말소리와 음성과학

Phonetics and Speech Sciences

한국음성학회지

http://dx.doi.org/10.13064/KSSS.2016.8.1.019

# The identification of Korean vowels /o/ and /u/ by native English speakers

Oh, Eunhae*

**Abstract**

The Korean high back vowels /o/ and /u/ have been reported to be in a state of near-merger especially among young female speakers. Along with cross-generational changes, the vowel position within a word has been reported to render different phonetic realization. The current study examines native English speakers' ability to attend to the phonetic cues that distinguish the two merging vowels and the positional effects (word-initial vs. word-final) on the identification accuracy. 28 two-syllable words containing /o/ or /u/ in either initial or final position were produced by native female Korean speakers. The CV part of each target word were excised and presented to six native English speakers. The results showed that although the identification accuracy was the lowest for /o/ in word- final position (41%), it increased up to 80% in word-initial position. The acoustic analyses of the target vowels showed that /o/ and /u/ were differentiated on the height dimension only in word-initial position, suggesting that English speakers may have perceived the distinctive F1 difference retained in the prominent position.

**Keywords:** non-native speech perception, Korean vowel merger, positional effects, L2 vowel identification

## 1. Introduction

Previous studies have reported several aspects of the on-going process of the /o/-/u/ merger in Korean (Kang, 1990; Yang, 1996; Chae, 2005; Sung, 2004, 2005; Han & Kang, 2013; Yoon, Schertz & Han, 2015). The raising of /o/ toward /u/, namely the coalescence of the vowel qualities resulted in reduction of phonetic contrast. However, the extent to how much two vowels are similarly produced has been shown to vary by age, gender, dialect, and syllable position. Han & Kang (2013) found a robust cross-generational difference in the production of the two vowels. Especially young female speakers of Seoul dialect were shown to make a significantly smaller perceptual distance between the two vowels while older female speakers and male speakers made a greater distinction. Similarly, Jang & Shin (2006) reported a generational difference in the vowel system of Seoul and Daegu dialect. The vowel changes observed in young female speakers of Daegu dialect exhibited a pattern similar to the vowels of Seoul dialect. These results are in line with the sociolinguistic view of women's high level of linguistic awareness and their progressive lead in sound change (Eckert, 1989; Labov, 2001).

Not only the degree but the direction of the merger differs across Korean dialects. Yoon et al. (2015) analyzed Korean vowels produced by speakers from eight different dialects of Korean and found notable dialectal variation. In particular, /o/ and /u/ in two of the central dialects (Chungbuk and Chungnam) were uniquely shown to have undergone a complete merger of vowel height (F1) which may account for their F2 sensitivity in differentiating the two vowels reported in Seong (2005). On the contrary, another central dialect (Gangwon) appeared to make contrast only along the F2 dimension. These differences suggest that the direction of the /o/ movement toward /u/ may vary across different dialects.

* Konkuk University, grace1111@konkuk.ac.kr
Received 2 February 2016; Revised 17 March 2016; Accepted 22 March 2016

The phonetic realization of the two vowels was also shown to differ by their position within a word. Kang (2015) found that /e/-/ɛ/ contrast (i.e., Korean vowel merger which is known to be complete for young speakers) was still observed in the speech of older male speakers but only in the word-initial position. As previously noted in Johnson & Martin (2003), vowels in word-final position are subject to greater reduction relative to vowels in word-initial position. At shorter duration, not only the information inherent in vowel spectral change is lost for perceptual contrasts but the compression of vowel space results in higher and more centralized vowels (Kang, 2015). If the identification of each vowel is more distinctively realized in word-initial position, we might expect non-native speakers of Korean to render relatively higher accuracy in discriminating /o/ and /u/ in initial syllables than non-initial syllables.

As for the non-native speakers, however, the effect of L1 on L2 speech perception should also be taken into account. That is, the phonetic parameters that they employ to distinguish the L1 vowel categories provide some clues as to which acoustic cues may be meaningful in L2 speech perception. In particular, the trajectory shape of formants and changes over the duration of the vowel has been reported to affect the identification of vowels (Nearey & Assmann, 1986; Strange, 1989) and the sensitivity to the dynamic acoustic properties appear to be acquired early in childhood (Nittrouer, 1992). Discriminating L2 vowels can be even more complicated when speakers of different languages give different perceptual weight to the acoustic properties of L2 vowels. In Nishi et al.'s perceptual assimilation studies (2008), native English listeners assimilated both Japanese long and short vowels to tense English vowels, suggesting that English listeners only attended to the spectral similarity but not temporal difference in distinguishing Japanese and English vowels. Chung, Kong & Weismer (2010) found that American English vowels are produced with greater spectral changes than Korean vowels. Considering that not only the magnitude but the direction of the spectral movements were found to be more salient and consistent for each vowel category in English, native English speakers may be more sensitive to the time-varying changes in the formant frequencies than Korean speakers when discriminating vowels.

The L2 speakers' difference in their use of perceptual cues to discriminate speech sounds has been observed in other features as well. Guion & Pederson (2007) examined the how the English and Japanese speakers differ in their use of perceptual cues to discriminate different tones in Mandarin. While the native Mandarin speakers mainly used the average F0 and slope of the F0 in perception of tones, English speaker relied on average F0 and extreme endpoints. Japanese speakers paid greater attention to average F0 and the high level tones. The results were interpreted to suggest that English and Japanese speakers attended to the cues as an influence of the intonation pattern and pitch accent in their native language respectively. In addition, English speakers with some knowledge of tone were investigated to see if they could acquire the ability to reweight and reconstruct optimal perceptual cues that native Mandarin speakers use for the tone perception. They found that the advanced learners of Mandarin were able to attend to the F0 slope when distinguishing different tones. The results indicate that language experience can redirect the learners' attention towards L2 phonetic contrasts.

It is assumed that greater number of high tense vowels in Korean than in English vowel inventory as well as the merger of Korean /o/ and /u/ present more challenge for English speakers to accurately discriminate the two vowels. In Yang (1996), (Southern) English /u/ was shown to be perceptually closer to Korean mid-central vowel /ɨ/ than /u/. Relatively low F2 values for Korean /o/ and /u/ provided sufficient distance from English high back vowels. However, the neighboring English diphthong /o/, particularly its final /ʊ/ portion of spectral movement, is likely to overlap with Korean /o/. If so, the expectation is that English speakers would identify Korean /o/ as an instance of /u/.

The current study was designed to investigate 1) whether native English speakers can distinguish the Korean merging vowels and if so, 2) whether there is a positional effect on the perceptual accuracy. Acoustic analyses of the vowels produced in word-initial and final position are expected show the relative contribution of the cues to the perceptual accuracy. Furthermore, on the assumption that the acoustic cues that were used to discriminate the L2 speech sounds may be represented in L2 production (Best, 1995; Flege, 1995; Sebastián-Gallés & Baus, 2005), the production of Korean vowels by the native English speakers was analyzed and compared the results with the perception experiment.

## 2. Method

### 2.1. Participants

Six native English speakers (three males and three females) from the western United States (California, Washington and Colorado) participated. The age of participants ranged from thirty five to forty three (Mean age = 39.5) and they had resided in Korea approximately three years at the time of testing (Mean length of residence = 2.6). They were all English instructors at a college located in Gyunggi Province. They all had some knowledge of Korean orthography and reported an average of 2.2. hours watching Korean TV or listening to Korean music on a daily basis.

### 2.2. Stimuli

Two vowel categories represented in twenty eight two-syllable real words, fourteen words exemplifying each vowel category, were produced by two female speakers of Seoul Korean. They were given some time to familiarize themselves with the stimuli. Each word randomly produced three times in a frame sentence 'ije ___ haseyo' ('now say ___') was recorded in DAT tape with a high quality microphone and were digitized at 22.05 kHz (16 bit). Seven out of the fourteen words were embedded in the first syllable and the other half in the second syllable. Some words contained the two vowels in both syllable positions (e.g., d**o**b**o** 'on foot'– d**u**b**u** 'bean curd'). The pre-consonantal contexts were matched across the syllable position (e.g., g**o**ri-g**u**ri, s**o**ri-s**u**ri, m**o**ja-m**u**jang vs. sug**o**-sug**u**, hos**o**-hos**u**, gom**o**-gom**u**), but word frequency inevitably varied across the stimuli. The CV portion of each target syllable were excised from the frame sentence and normalized to 50% peak intensity before they were randomly presented to six native English speakers.

### 2.3. Procedure

A two-alternative forced-choice identification task was given to the participants. They were asked to press either 1 for /o/ or 3 for /u/ presented on a computer monitor with Korean orthography. Two acoustic features of the given speech sounds, the duration and

formant values, were analyzed. With the duration of the vowel selected in Praat, first and second formant measurements were taken at the 25%, 50% and 75% points (marked as _1, _2, _3 in Figure 2 and 3) of the vocalic portion. The vowels were measured from the onset of the first periodic wave to the offset of the last one observed in both the waveform and the spectrographic display. A total of 168 speech sounds (28 words x 2 speakers x 3 repetitions) were evaluated. Because the durations and formant frequencies showed little differences across the three repetitions, mean values were used in the analyses.

In order to further examine the English speaking participants' production of the Korean vowels, two male participants who were able to read Korean were asked to read the target words in a frame sentence three times in a random order. Korean words with matched consonantal context included the two vowels in the first syllable such as 'gori (hook), gugəl (beg), guri (copper)'. Different target words were elicited because they needed to be easy to pronounce. Although the target words were different between the native Korean and English speakers, words with the same pre-consonantal context were selectively compared for the analysis. First and second formant frequencies were measured at the temporal midpoint of each vowel.

## 3. Results

The results indicated that the overall identification accuracy was significantly lower for /o/ compared to /u/. The accuracy was the lowest for /o/ produced in word-final position (42.8%). As shown in Figure 1, however, more than 80% of the /o/ vowel produced in the initial position was accurately identified by English speakers.

The proportions of the accurately identified tokens for each intended vowel category were arcsine transformed to ensure normality and submitted to an ANOVA. There was main effects of vowel [$F(1,88) = 97.023$, $p < .001$], position [$F(1,88) = 28.087$, $p < .001$] as well as a significant interaction between vowel and syllable position [$F(1, 88) = 5.131$, $p = .026$]. The positional effect was significant only for /o/ [$F(1,44) = 47.089$, $p < .001$]. As clearly illustrated in the figure below, the number of accurately identified /o/ in initial position roughly doubled that in final position.

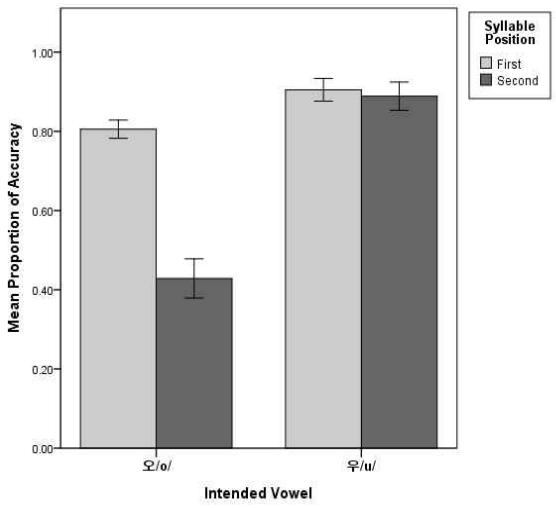


**Figure 1.** Normalized proportion of accurately identified vowels embedded in word-initial and final position is shown.

To further examine the magnitude of the two variables' contribution to the accurate identification, logistic regression analyses were conducted. For each intended vowel, the independent variable of syllable position was entered as predictor of the dependent variable of normalized proportions of accurately identified tokens.

**Table 1.** Results from the logistic regression

| Vowel | Variable | B (S.E) | Wald (df=1) | Odd Ratio (Exp(B)) |
|---|---|---|---|---|
| o | Position | 1.71(.204) | 70.319 | *5.524 |
| | Constant | -.288(.127) | 5.108 | .750 |
| u | Position | .172(.294) | .342 | 1.187 |
| | Constant | 2.079(.200) | 7.621 | .822 |

Table 2 indicates that the effects of variable were significant for /o/. Particularly, odds ratio *(Exp (B))* of syllabic position suggested that the odds of a /o/ receiving accurate response was 5.5 times greater it was in initial than final position.

The vowel duration and formant values of each vowel embedded in different syllable positions were compared to examine the acoustic characteristics which might have led to the difference in accuracy. The effect of syllable position on the mean duration of the two vowels was analyzed using ANOVAs with a repeated measures design. The results showed no significant main effect of vowel or the vowel and position interaction, suggesting that vowel duration did not significantly vary by position.

An ANOVA was conducted to determine whether a difference in formant frequencies existed between vowels produced. The two vowels embedded in each syllable position were separately analyzed. First, the results indicated all three points of the F1 [$F(3,18) = 12.332$, $p < .001$] as well as the F2 values [$F(3,18) = 14.788$, $p < .001$] differed between the two vowels in initial position. As shown in Figure 2, /o/ was produced with higher F1 and lower F2 values than /u/ and the differences remained distinctive over time.

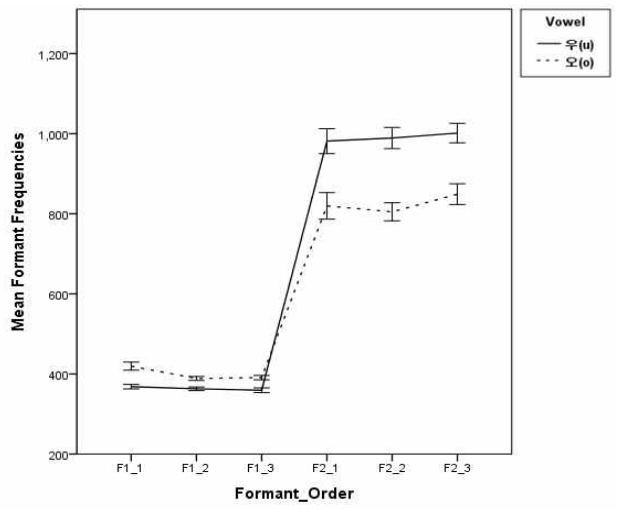


**Figure 2.** Mean F1 and F2 frequencies of Korean vowel /o/ and /u/ produced in initial position are shown.

As for the final position, the F1 values showed no effect of vowel [$F(3, 18) = 1.124$, $p = 366$]. None of the three points differed between the two vowels. However, the F2 values significantly differed by vowel [$F(3, 18) = 12.482$, $p < 001$]. Although there is a slight drop in the F2 values for /o/, there was no significant difference across the three points for either /o/ or /u/.

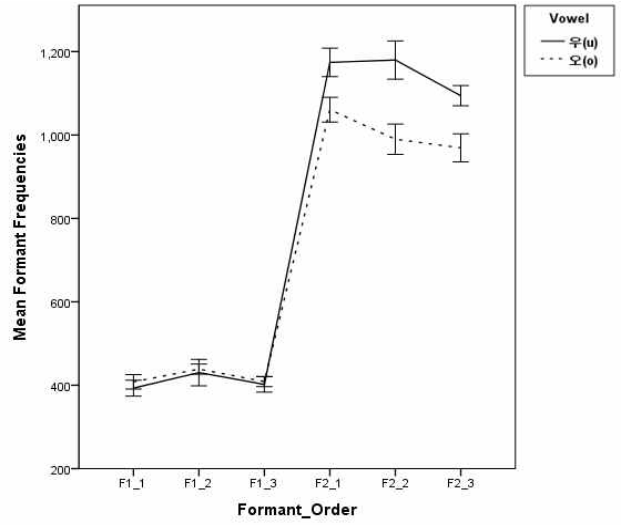


**Figure 3.** Mean F1 and F2 frequencies of Korean vowel /o/ and /u/ produced in final position are shown.

The post-hoc analysis on each vowel showed a significant increase in the mean F1 values (i.e., a lower tongue body position) for /o/ and /u/ in final position and greater lowering of /u/ gave rise to the F1 merger of the two vowels. Similarly, the mean F2 values for both /o/ and /u/ were significantly increased in final position ($p < .001$). What can be inferred from Figure 2 and 3 is that the English speakers may have perceived the distinctive F1 difference retained in initial position.

Linear mixed effects modelling was used to further assess whether the difference in the F1 and F2 frequencies produced in the word-initial and final position influenced the perceptual accuracy. The midpoint of the F1 and F2 frequencies were submitted as dependent variables and listeners' responses were taken as an independent variable. As for the vowels in word-initial position, there was a significant effect of listeners' responses on both F1 [$F(1,484) = 12.096$, $p < .001$] and F2 frequencies [$F(1,484) = 69.579$, $p < .001$] while the vowels in word-final position showed an effect on only F2 frequencies [$F(1,484) = 6.527$, $p = .011$]. The results are interpreted to indicate that listeners were likely to exploit both F1 and F2 frequencies in word-initial position whereas only the F2 frequencies were used as an important cue to distinguish the two vowels in word-final position.

On the assumption that perception precedes production, English speaking participants' production of the two vowels embedded in the first syllable position was examined. As shown in Figure 4, some of the Korean vowels elicited from the two Korean male speakers (marked with a dotted line) were put together with Korean vowels produced by two of the English-speaking participants. The results showed a significant main effect of group [$F(2,9) = 5.306$, $p = .030$] as well as a significant vowel by group interaction [$F(2,9) = 9.220$, $p = .007$]. When each vowel was separately examined, the effect of group was significant on the F1 values [$F(1,10) = 14.310$, $p = .004$] but not on the F2 values [$F(1,10) = 3.514$, $p = .090$] for /o/. As for /u/, neither F1 [$F(1,10 = 2.021$, $p = .186$] nor F2 values [$F(1,10) = 2.041$, $p = .184$] differed by group. Namely, the native English speakers produced /o/ with higher F1 than the Korean speakers in word-initial position.

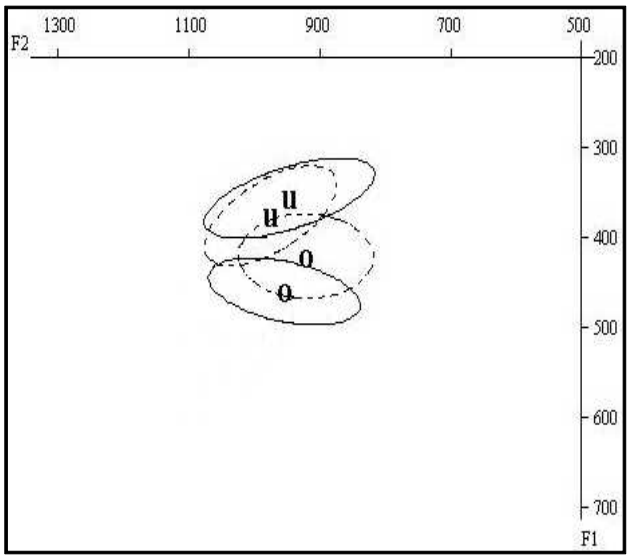


**Figure 4.** Mean F1 and F2 frequencies of the Korean vowels /o/ and /u/ produced by two native Korean and English male speakers are shown (two English speakers' vowels are marked with a dotted line).

## 4. Discussion

The current study showed that the native Korean speakers made a clear F1 and F2 distinction in word-initial position and the merger of /o/ and /u/ was observed in word-final position but only on the height dimension. Due to the significant positional effects on vowel realization, the native English speakers showed difficulty discriminating the two L2 vowel categories significantly more so than they were positioned word-finally.

The degree of identification accuracy also depends on whether the L2 vowels are represented as distinctive categories in the native vowel inventory. According to the Native Language Magnet model (Iverson & Kuhl, 1995) and the Perceptual Assimilation Model (Best, 1995, 2007), L2 learners show difficulty in the perception of L2 sound contrasts that are not distinctive in their native language. That is, L2 vowel categories are likely to be mapped onto native categories based on their perceived similarity. Despite the differences in the vowel trajectory of Korean and English /u/ (Yang, 2010; Chung et al., 2010), their similar target F1 and F2 values of English and Korean /u/ can be predictive of accurate vowel perception for Korean /u/. On the one hand, the perceptual similarity between Korean and English /u/ may account for the overall higher accuracy for Korean /u/ regardless of syllable position. In final position, however, /u/ showed a significant increase in the F1 and this lowering of the high vowels in weaker position is in agreement with previous observation of the position-conditioned variation shown in Kang (2015). Namely, high vowels in final position were realized with a lower and more central tongue body position than those occurring in initially position due to the compression of the

vertical vowel space. When the F1 difference was obscured, /o/ was also perceived as /u/ 60 % of the time, suggesting that the two Korean vowels were assimilated to a single perceptual representation in final position.

On the other hand, the twofold increase in the ability to accurately identify /o/ in initial position (80.5%) is in line with previous studies showing retention of contrast among vowel mergers solely in word-initial position (Jung 2002; Kang, 2015). The positional prominence of initial position accounts for the English speakers' ability to identify the L2 vowel categories that are not distinctive in the native language.

It should be noted that although the F1 difference was more evident word-initially, it is not clear as to whether the English speakers were able to attend to the cue to facilitate target identification. Assuming that there is a strong correlation between L2 perception and production and that the perceptual cues listeners use for a phonological contrast are manifested in their productions of the contrast (Beddor, 2015; Kwon, 2015), the English speakers' production of the two vowels was examined. As shown in Figure 4, /o/ and /u/ differed only in F1 values which may suggest that the English speakers' sensitivity to F1 difference between the two vowels is reflected in production as well.

English speakers' greater use of F1 frequencies in discriminating vowels have been reported in previous studies (Kewley-Port & Watson, 1994; Assmann & Katz, 2000). In Assmann & Katz (2000), 12 vowels produced in /hVd/ context by 20 adults were fully and partially synthesized and presented to English speaking adult listeners for identification and discrimination tasks. Among the synthesized version, *Flat* F1 (i.e., the center frequency of the first formant was held constant) showed significant declines in vowel identification for /o/ while there was little effect of *Flat* F2. If the identification accuracy for /o/ in English was substantially greater when the F1 information was fully provided, the implication is that the F1 difference shown in initial position may have offered benefits for identifying the two Korean vowels. Consistent with previous studies, the results L2 speakers appear to give different weight to the acoustic cues for discriminating L2 vowels due to the L1 influence.

The results, however, should be interpreted with caution as there were only two English speakers and their English production was not elicited. Assuming that there is a strong correlation between L2 perception and production and that the perceptual cues listeners use for a phonological contrast are manifested in their productions of the contrast (Beddor, 2015; Kwon, 2105), the phonological representations of the L1 and L2 vowel systems can provide more accurate view of the L1 influence on the perception of L2 vowels. Moreover, without further information on F0, F3 frequency and spectral envelope which were shown to provide additional cues for vowel identification (Ryalls & Lieberman, 1982), it is beyond the scope of this study to determine the full degree of acoustic similarity between the vowels.

## References


Assmann, P. & Katz, W. (2000). Time-varying spectral change in the vowels of children and adults. *Journal of Acoustical Society of America*,108, 1856-1866.

Beddor, P. (2015). The Relation between Language Users' Perception and Production Repertoires. *Proceedings of the 18th ICPhS*.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues* (pp. 171 – 204). Timonium, MD: York Press.

Best, C. T. & Tyler, M. (2007). Nonnative and second language speech perception: Commonalities and complementaries. In O.-S. Bohn & M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege* (pp. 13-34). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Chae, S. (1999). The core-periphery structure in the Korean lexicon reflected in a phonological variation and change. *Studies in Phonetics, Phonology and Morphology*, 5, 217-236.

Chung, H., Gong, E., & Weismer, G. (2010). Vowel formant trajectory patterns for shared vowels of American English and Korean, *Phonetics and Speech Sciences,* 2, 67-74.

Eckert, P. (1989). The whole woman: Sex and gender differences in variation. *Language Variation and Change*, 1, 245-268.

Flege, J. (1995). Second language speech learning theory, findings, and problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 233-277). Timonium, MD: York Press.

Guion, S. G. & Pederson, E. (2007). Investigating the role of attention in phonetic learning. In O.-S. Bohn & M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege* (pp. 57-78). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Han, J. & Kang, H. (2013). Cross-generational Change of /o/ and /u/ in Seoul Korean I. *Phonetics and Speech Sciences*, 5, 25-31.

Iverson, P. & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America*, 97, 553 – 562.

Jang, H. & Shin, J. (2006). An acoustic study on the generational difference of the monophthongs in the Daegu Dialect. *Malsori*, 57, 15-30. (장혜진·신지영 (2006). 대구 방언 단모음의 세대 간 차이에 대한 음향 음성학적 연구. *말소리*, 57, 15-30.)

Johnson, K. & Martin, J. (2001). Acoustic vowel reduction in Greek: Effects of distinctive length and position in the word. *Phonetica*, 58, 81-102.

Jung, M. (2002). Sound change of language used in broadcasting, *Kwukehak*, 39, 221-249. (정명숙 (2002). 방송 언어에 나타난 말소리의 사적 변천, *국어학*, 39, 221 – 249.)

Kang, Y. (2015). A corpus-based study of positional variation in Seoul Korean Vowels. In M. Kenstowicz, T. Levin, & R. Masuda (Eds.), *Japanese/Korean Linguistics*, 23. CSLI Publications.

Kewley-Port, D. & Watson, C. S. (1994). Formant-frequency discrimination for isolated English vowels, *Journal of Acoustical Society of America*, 95, 485-496.

Kwon, H. (2015). *Cue primacy and spontaneous imitation: is imitation phonetic or phonological?* Ph.D. Dissertation, University of Michigan, Ann Arbor.

Labov, W. (2001). *Principles of Linguistic Change*, *Vol. 2: Social Factors*. Malden, MA: Blackwell Publishers Inc.

Nearey, T. M. & Assmann, P. F. (1986). Modeling the role of inherent spectral change in vowel identification. *Journal of the Acoustical Society of America*, 80, 1297-1308.

Nishi, K., Strange, W., Akahane-Yamada, R., Kubo, R., & Trent-Brown, S. A. (2008). Acoustic and perceptual similarity of

Japanese and American English vowels. *Journal of Acoustical Society of America*, 124, 576 – 588.
Nittrouer, S. (1992). Age-related differences in perceptual effects of formant transitions within syllables and across syllable boundaries. *Journal of Phonetics*, 20, 351 – 382.
Ryalls, J. H., & Lieberman, P. (1982). Fundamental frequency and vowel perception. *Journal of the Acoustical Society of America*, 72, 1631-1634.
Sebastián-Gallés, N. & Baus, C. (2005). On the relationship between perception and production in L2 categories. In A. Cutler (Ed.), *Twenty-First Century Psycholinguistics: Four Cornerstones* (pp. 279 – 292). New York: Erlbaum.
Seong, C. (2005). A formant analysis of the Korean monophthongs of the college students speaking Chungnam dialect. *Eoneohag*, 43, 189-213. (성철재 (2005). 충남지역 대학생들의 한국어 단모음 포먼트 분석. *언어학*, 43, 189-213.)
Strange, W. (1989). Dynamic specification of coarticulated vowels spoken in sentence context. *Journal of the Acoustical Society of America*, 85, 2135-2153.
Yang, B. (1996). A comparative study of American English and Korean vowels produced by male and female speakers. *Journal of Phonetics*, 24, 245-261.
Yang, B. (1996). Perceptual contrast in the Korean and English vowel system normalized. *ICSLP-1996,* 114-117.
Yang, B. (2010). Formant trajectories of English high tense and lax vowel produced by Korean and American speakers, *Korean Journal of Linguistics*, 35, 407-421.
Yoon, T., Kang, Y., Han, S., Maeng, H., Lee, J., & Kim, K. (2015). A corpus-based approach to dialectal variation in Korean vowels. *Proceedings of the 18th ICPhS.*

• **Oh, Eunhae**
Department of English Language and Literature
Konkuk University
120 Neungdong-ro, Gwangjin-gu
Seoul 05029, Korea
Tel: 02-450-3046
Email: grace1111@konkuk.ac.kr