Korean Journal of English Language and Linguistics
Vol. 20, March 2020, 20−41
10.15738/kjell.20..202003.20

Production and Perception of Phonetic Variations across English Word Boundaries by Korean EFL Learners

Hye−Young Um
(Myongji University)

Um, Hye−Young. 2020. Production and perception of phonetic variations across English word boundaries by Korean EFL learners. *Korean Journal of English Language and Linguistics* 20, 20−41. The present study aimed to investigate whether and how Korean L2 English learners' verbal productions of near−homophonous phrases (e.g., *keeps parking* vs. *keep sparking*; *seen either* vs. *see neither*) are distinctively perceived by native speakers of English. It also examined whether and how Korean learners correctly perceive the native speaker's production of those near−homophonous phrases. For both correct production and perception, utilizing acoustic−phonetic information such as aspiration and glottal stop is necessary to signal and identify word boundary within a phrase. Twenty Korean college EFL learners participated in the production and perception experiments. The results showed: 1) Korean learners' production of the phrases that necessitate a positive phonetic cue (e.g., *keeps parking*) was more difficult for native English speakers to correctly perceive; 2) Korean learners' production of aspiration type was more difficult to be correctly perceived than the other types; and 3) Korean learners had difficulty in perceiving the aspiration type phrases and the phrases that do not have positive phonetic cues, although they were generally very good at the perception task. Possible sources for Korean learners' difficulty in producing positive phonetic cues were discussed.

Keywords: word boundaries, segmentation, phonetic variations, phonetic cues, production, perception, aspiration, glottal stop, second language

1. Introduction

To understand the target language speech successfully, second language learners need to be able to use acoustic phonetic cues that are relevant in segmenting speech into words. Also in order to produce the target language successfully and make the target language speakers understand what they say as intended, they need to be able to produce allophonic variations appropriately in the stream of speech. For example, in English word *speech* /p/ is not aspirated, but even if it is produced as aspirated, it

would not cause much confusion being understood as a foreign accent. However, in the phrase like *lay speech* it would result in miscommunication if /p/ is produced as aspirated. The phrase may be perceived as *lace peach* instead of *lay speech* when no situational context for the use of the phrase is assumed. Similarly, in listening to the phrases like *a nice man* or *an ice man*, native speakers of English distinguish the two phrases by relying heavily on the presence or absence of glottal stop inserted before [ɑɪs] (Nakatani and Dukes 1977). That is, they use the presence or absence of glottal stop/glottalization as a perceptual cue when listening to the phrase. Because native speakers of English produce phrases that contain acoustic-phonetic information needed for word segmentation, and the native listeners of English use that information to correctly segment phrases into words, it may be worthwhile to see whether L2 learners of English can produce and use the same acoustic-phonetic information for their production and perception of English phrases.

There have been several studies on L2 learners' ability to utilize acoustic-phonetic cues that native speakers of English use in perceiving near-homophonous phrases (Altenberg 2005, Ito and Strange 2009, Shoemaker 2014, Um 2006). These studies investigated L2 English learners' perception of near-homophonous phrase pairs in which the sequence of phonemes is identical, but the location of word boundary is different (e.g., *keeps parking* vs. *keep sparking*). However, there have been little research on how and whether L2 learners produce appropriate phonetic variations involved so that native speakers of English can perceive what they intended to say.

The present study aims to investigate how Korean learners of English produce phonetic variations such as aspiration/nonaspiration and presence/absence of glottal stops across word boundaries. It will also investigate whether these learners successfully identify word boundaries in their perception, and see whether there is a correlation between their production and perception of these phonetic variations. In the following sections, previous studies on this issue will be briefly reviewed, and then the methods and results of present experimental studies will be reported.

2. Previous Research

In order to segment the incoming speech into words, various kinds of information such as lexico-semantic information, syntactic/morphological information, sentential context, prosodic rhythmic cues and allophonic acoustic-phonetic information are used

by listeners (recited from Ito and Strange 2009). In the study on the use of acoustic-phonetic segmentation cues by native speakers of English, Nakatani and Dukes (1977) found that the occurrence of allophonic variations of the same phoneme can signal word boundary. According to them, strong cues for word juncture in English are glottal stops or laryngealization at the onset of word-initial vowels, aspiration of voiceless stops and distinct /l/ and /r/ allophones. Therefore, in the ambiguous phrases such as *keepstalking*, if aspiration is heard on the stop /t/, then the listener will segment it as *keeps talking*. If not, it will be perceived as *keep stalking*. In the phrase such as *seen either*, glottal stop and/or laryngealization (creaky voice), often inserted before word-initial vowels in English is a strong cue for boundary perception, distinguishing the phrase from *see neither*. In the phrase such as *sail east*, the allophonic variation of /l/ (velarized) and the presence of glottal stop make it distinguishable from *say least*.

Starting from Altenberg (2005), there have been some studies that reported the abilities of second language learners of English to use these acoustic-phonetic cues in segmenting two or three-word near-homophonous English phrases into words (Um 2006, Yun 2009, Chang and Park 2011 on Korean learners of English, Ito and Strange 2009 on Japanese, Shoemaker 2014 on French speakers). All of these perceptual studies showed that the learners of English as a second language were not successful in fully using the kinds of acoustic-phonetic cues available to native English speakers.

Altenberg (2005) examined acoustic cues to word juncture in Spanish learners of English using an adaptation of Nakatani and Dukes's (1977) design that uses pairs of phrases in which the sequence of segmental phonemes is identical, but the location of word boundary is different. She found that Spanish speakers at both intermediate and advanced levels of English proficiency performed above chance in the identification of potentially ambiguous stimuli (76%), but that they performed significantly worse than native speakers. Stimuli that included glottal stop were identified with more precision than those that included aspiration (88.4% vs. 58.5%). She attributed this discrepancy primarily to L1 transfer, noting that glottal stops occur often in Spanish language, whereas aspiration is rare. She also explored the possibility that markedness could play a role in that the insertion of a glottal stop before a word-initial vowel is unmarked.

Extending the study of Spanish L2 learners by Altenberg, Ito and Strange (2009) conducted a similar perception experiment with Japanese speakers, which yielded generally similar results to those reported by Altenberg for Spanish learners. Similarly

to Spanish speakers, Japanese subjects' performance on aspiration stimuli was significantly poorer than that on glottal stop stimuli (73.1% vs. 91.3%). They found that their subjects' performance correlated with length of residence and suggested that aspiration cues took more immersion experience to learn than glottal stop cues.

Shoemaker (2014) is on the perception study with French learners of English. From the experiment conducted using the same recordings used in Ito and Strange (2009), she found that, similarly to Spanish and Japanese subjects, French speakers' performance was significantly worse than that of English speakers and that her participants were also more sensitive to the presence of glottal stop than aspiration. She also found that there was a clear effect of language exposure in that third-year participants performed significantly better than first-year participants. In her discussion of the subjects' sensitivity to aspiration, she suggested that the reason that Spanish and French speaking learners were worse than Japanese speakers was that aspiration of initial voiceless plosives is used more systematically in Japanese than in Spanish and French.

Contrary to the expectation that the learners of English whose native language uses aspiration contrastively are more sensitive to aspiration in identifying word boundaries, Um (2006) found that Korean learners also had more difficulty perceiving aspiration stimuli than glottal stimuli. Subsequent studies with Korean learners (Chang and Park 2011, Yun 2009) also reported similar results. This might indicate the universal perceptual salience of glottal stop cue as Shoemaker (2014) suggested.

Um (2006) additionally found that Korean learners were better at segmenting stimuli with a positive cue than a negative one both in aspiration (63.67% vs. 48.57) and glottal stimuli (90.71% vs. 75%). It was suggested that the presence of a cue is perceptually more salient than its absence. The better performance on the positive stimuli was also observed in the results of Yun (2009) and Chang and Park (2011).

While all the previous studies on this issue have looked into the L2 perception of word boundary cues, there was no study that I know of that has investigated whether learners of English produce phonetic variations such as presence/absence of aspiration and glottal stops across word boundaries correctly enough for native English speakers to understand their production. In the next sections, I will present the results of the production experiment along with those of the perception experiment and see if there is a correlation between the production and perception.

3. Method

3.1 Subjects

A total of 20 Korean students, 11 female and 9 male at a university in Korea, ranging in age from 20 to 26 participated in the experiments. In analyzing the results of the experiments, they were divided into two groups — advanced and intermediate. Advanced group consists of 10 students (4 male, 6 female) whose TOEIC scores range from 850 to 990 (average 900). Intermediate group consists of 10 students (5 male, 5 female) whose TOEIC scores range from 500 to 650 (average 585).[1]

3.2 Materials

I used the similar materials used by Altenberg (2005) and Ito and Strange (2009), but adjusted them slightly for the following reasons: 1) to exclude unfamiliar words and proper nouns; and 2) to reorganize double cue stimuli (e.g., *say least* and *sail east* were categorized as the glottal stop stimulus type, but I categorized them as double type because the allophones of liquid /l/ are also involved as well as the presence or absence of glottal stop in demarcating the word boundary). Total of 23 pairs (46 phrases) were used in both the perception and production experiments.

The stimuli are divided into three types: Nine aspiration stimulus pairs (18 phrases) such as *lace peach* and *lay speech*, in which the presence or absence of aspiration of the stops in the second word's onset position provides strong segmentation cue; Eight glottal stimulus pairs (16 phrases) such as *seen either* and *see neither*, where the presence or absence of glottal stop which is inserted before word-initial vowels and/or creaky voice distinguishes the two phrases from each other; Six double stimulus pairs (12 phrases) such as *gray pin* and *grape in* in which two different kinds of phonetic variations signal word boundary.

Aspiration stimuli are further divided into three format groups which differ in the amount of aspiration: 1) the one in which /s/+C comes after a vowel (e.g., *lace peach* vs. *lay speech*); 2) the one in which /s/+C comes after a consonant (e.g., *keeps parking* vs. *keep sparking*); and 3) the one in which /s/+C comes after a consonant

[1] TOEIC scores have been used to divide EFL learners' proficiency levels in many of the previous L2 studies. However, it may be fair to mention here that TOEIC score doesn't necessarily reflect the subjects' proficiency in all aspects.

and is followed by another consonant /r/ (e.g., *cook's prints* vs. *cook sprints*). Glottal stimuli are broken down into two formats: 1) the one that has a nasal consonant at the word boundary (e.g., *seen either* vs. *see neither*); and 2) the one that has an obstruent at the word boundary (e.g., *wife ill* vs. *why fill*). Double stimuli are also divided into two format groups: 1) the one in which the phonetic characteristics of liquid /l/, (i.e., dark or clear /l/) and the presence or absence of glottal stop play a role in deciding the word boundary (e.g., *I'll earn* vs. *I learn*)[2]; and 2) the one in which both the presence or absence of aspiration and that of glottal stop play a role in signaling the word boundary (e.g., *gray pin* vs. *grape in*).

In presenting the results of the experiment, the stimuli in which the presence of aspiration or glottal stop serves as a boundary cue will be referred to as positive stimuli (e.g., *lace peach* and *seen either*) and those in which aspiration or glottal stop is absent will be referred to as negative stimuli (e.g., *lay speech* and *see neither*). The second format of double stimuli is not divided into positive or negative subcategories. However, the first format of double stimuli will be divided into positive/negative subcategories for convenience, because the phrases in this type can be segmented into words depending both on whether velarization of /l/ applies and on the presence or absence of glottal stop.

Table 1 shows the target items used in this study.

[2] Altenberg (2005), Ito and Strange (2009) and the subsequent studies on this issue categorized this type of stimuli as glottal stop stimuli. However, this type of stimuli was grouped as Double stimuli in this paper because two kinds of strong cues, that is, the distinct allophonic variations of 'l' and the presence or absence of glottal stop/creaky voice are involved.

Table 1. Stimulus Pairs for Production and Perception Experiments

	Format	+Asp (positive)	−Asp (Negative)
Aspiration	VsC	lace peach lace table lace car	lay speech lay stable lay scar
	CsC	keeps parking keeps talking chief's cool	keep sparking keep stalking chief school
	CsCC	cook's prints cook's truck cook's cream	cook sprints cook struck cook scream
		+Glottal (positive)	−Glottal (Negative)
Glottal	VNV	seen either an ice man team at claim annual	see neither a nice man tea mat clay manual
	VOV	wife ill loaf ate grave at weave end	why fill low fate gray vat we vend
Double[3]		Dark 'l' (+velarization)[4]/ +Glottal	Clear 'l' (−velarization)/ −Glottal
	VLV	I'll earn sail east seal of	I learn say least see love
		+Asp/ −Glottal	−Asp/ +Glottal
	VCV	gray pin lie told lie cold	grape in light old like old

[3] Previous studies including Altenberg (2005) and Ito and Strange (2009) named this type of stimuli as 'Double cue stimuli' because they used these materials for their perception studies. In this paper this type of stimulus pairs will be referred to as 'Double' without the word 'cue' since these materials were used both in perception and production experiments.

[4] Strictly speaking, unlike aspiration and glottal stop, velarization does not mean the presence/absence of a phonetic cue or phonetic characteristics. Rather it is a matter of phonological rule being applied. However, I used the term +/− velarization for convenience because the stimuli involving velarization can be said to "have" the velarized characteristics of /l/.

3.3 Procedure and Analysis

For the perception experiment by Korean participants, 46 stimuli items (23 pairs) were prepared, recorded by a male native speaker of English who has been teaching English at a university in Korea. Each stimulus was placed within the carrier sentence "Say _____ again", and the stimuli were shuffled using Excel. The recorded speech was confirmed as natural by another native speaker of English who also has been teaching English in Korea.

The production task was given first. In this task the Korean participants were asked to read the target items in a carrier sentence "Say _____ again". They were given enough time to practice and were directed to read the items as naturally as possible with the same rate and stress on each word. They were also instructed to read the items again in case they think they produced the stimuli incorrectly. The 46 target items (23 pairs) were randomized using Excel program so that they were in different order from the perception task. The recordings were made individually using a microphone in a quiet room and each participant's recording was saved as a wave file.

Three native speakers of English who are all instructors at a university in Korea, all male, in their late 30s and 40s in age, worked as judges about what the Korean EFL speakers said. They listened to Korean subjects' production of target items through headsets and were asked to choose what they heard from two written alternatives (e.g, ① keeps parking ② keep sparking). Each native speaker of English listened to all 20 students' production files and the average score of the three native speakers' perception on each target item was used as the subject's production score on each target item.[5]

Korean participants were given perception task a few days after they had finished their production task. Participants listened to the stimuli through headsets and were asked to choose what they have heard from the two written alternatives. They were instructed to pay attention only to the sound, but not to the meaning. Perception experiment was conducted without native English controls because all previous studies reported that native speakers of English were successful in utilizing word boundary cues.

[5] Three native English speakers have been teaching English in Korea for an average of 10 years, and admittedly, their long-term exposure to English spoken by Korean learners might affect their perception of target items. Two of the native English speakers are from the U.S. and the other is from Canada. The inter-rater reliability was high (ICC = .82).

4. Results

4.1 Production: Perceived by Native English Speakers

The mean percentages of Korean learners' (NKs hereafter) production items that were correctly perceived by native English speakers (NEs hereafter) are presented in Table 2.

Table 2. Mean Percentage of Correctly Perceived NK Production by Stimulus Type and Proficiency Group

Stimulus Type	Advanced Group Mean (N = 10)	Intermediate Group Mean (N = 10)	Total Mean
Aspiration	58.52	60.93	59.72
Glottal	82.50	78.75	80.63
Double	90.00	85.00	87.50
Total	76.17	73.78	74.98

Overall, the NEs correctly perceived NKs' intended production at about 75% rate. The mean of the advanced group was 76.17% and that of the intermediate group was 73.78%, and no significant difference was found between the two groups' means (t = .622, p > .05). The result indicates that advanced and intermediate level learners were not different in the rate of the production that is correctly perceived by NEs. An ANOVA analysis confirmed that there was no main effect of English proficiency group (F = 1.063, p > .05), but there was a significant main effect of stimulus type (F = 66.337, p < .001). Post-hoc analysis indicated that means of the three types were significantly different from one another: Aspiration type productions were least correctly perceived by NEs (59.72%), followed by the glottal stop type productions (80.67%) and the double type productions (87.5%). As there was no interaction effect between the stimulus type and English proficiency (F = 1.170, p > .05), the results indicate that, regardless of their English proficiency level, NKs have the most difficulty in correctly pronouncing the aspiration type items whereas they have the least difficulty in correctly pronouncing the double type items.

Analyses were also made to see the effects of the positiveness and the stimulus format within each stimulus type. Table 3 presents the mean percentages of correctly perceived NK production of three different aspiration format items by positiveness.

Table 3. Mean Percentage of Correctly Perceived NK Production of Aspiration Type by Format and Positive/negativeness

Aspiration Format	Positive	Negative	Total Mean
VsC	44.42	76.11	60.27
CsC	47.22	72.22	59.72
CsCC	51.11	67.22	59.16
Total	47.58	71.85	59.72

Two-way repeated measure ANOVA revealed that the differences among the three aspiration formats were not significant (F = .071, p > .05). However, there was a significant main effect of positiveness of aspiration stimuli (F = 25.357, p < .001). No interaction effect was shown between the format and positiveness (F = 3.568, p > .05). These results indicate that, to NEs, NKs' production of the positive aspiration stimuli (47.58%) was significantly more difficult to perceive than negative aspiration stimuli (71.85%) was, across all three formats.

Table 4 shows the mean percentages of correctly perceived NKs' production of the glottal type by each format and positiveness.

Table 4. Mean Percentage of Correctly Perceived NK Production of Glottal Type by Format and Positive/negativeness

Glottal Format	Positive	Negative	Total Mean
VNV	76.67	90.42	83.54
VOV	71.67	83.75	77.71
Total	74.17	87.08	80.63

Because the scores of negative items are not normally distributed, a nonparametric Wilcoxon signed-ranks test was used.[6] For the glottal type, the difference between the two formats was not significant (z = −1.654, p > .05), but there was a significant effect of positiveness (z = −3.270, p < .001). Again, across the two formats, the NEs better perceived NKs' production of negative stimuli than the positive stimuli, although the difference between the NEs' perception of the positive (74.17%) and negative (87.08%) stimuli in the glottal type (approximately 13%) was not as high as that in the aspiration type (approximately 24%).

[6] Because of the small number of samples in this study, Shapiro-Wilk test was used to assess the normality assumption for all the data used in this study. Normal distribution could not be assumed for NK production data of glottal type negative items (for the both formats), double type items (for the both formats), and for most data used for NK perception analyses.

Table 5 presents the results of the double type production.

Table 5. Mean Percentage of Correctly Perceived NK Production of Double Type by Format and Positive/negativeness

Double Format	Positive	Negative	Total Mean
VLV	87.22	91.11	89.17
VCV	85.83		85.83
Total			87.50

Overall, NEs correctly perceived NKs' production of both formats of double type at a very high rate (89.17 for the format VLV and 85.83 for the format VCV). Wilcoxon signed-ranks test found no significant difference between the two formats ($z = -.786$, $p > .05$). Also, there was no difference between positive and negative stimuli in the format VLV ($z = -.630$, $p > .05$). Comparison for positiveness was not made for the format VCV items because only the format VLV items could be distinguished in terms of positiveness. The results indicate that NEs have little difficulty in correctly perceiving the NKs' production of the double type items irrespective of the format and positiveness.

To sum, the results of the production experiment showed that Korean subjects' English proficiency (as judged by their TOEIC scores) was not significantly associated with their ability to make verbal production intelligible to NEs. However, their performance differs significantly by stimulus types and positiveness. Their production of aspiration type was most difficult for NEs to perceive than other types, and the difficulty was more severe with the items with positive aspiration cue (such as *keeps parking*) than the ones without it (such as *keep sparking*). The glottal type was less difficult to perceive than the aspiration type, but NKs' production of the items with positive glottal cue (such as *seen either*) was more difficult for NEs to perceive than the ones without it (such as *see neither*). NEs perceived NKs' production of the double type most easily. No effect of positiveness was found for this type.

4.2 Perception

Table 6 reports the mean percentages of correct responses for each stimulus type by the two proficiency groups in NK perception test.

Table 6. Mean Percentage of Correct NK Perception by Stimulus Type and Proficiency Group

Stimulus type	Advanced Group Mean (N=10)	Intermediate Group Mean (N=10)	Total Mean
Aspiration	83.89	77.78	80.83
Glottal stop	96.25	96.87	96.56
Double	93.33	96.67	95.00
Total	91.16	90.44	90.80

Overall, NKs performed very well on the perception task: In total, they were able to correctly identify target items at a rate of around 90%, with little difference between advanced (91.16%) and intermediate (90.44%) groups. Because of skewed distributions seen in the scores of the perception test, a nonparametric Mann-Whitney test was used to compare the advanced and intermediate groups and nonparametric Friedman's tests and Wilcoxon signed-ranks tests were used to examine differences across three stimulus types. Mann-Whitney test confirmed that there was no significant difference between the advanced and intermediate groups in total (u = 8277.0, p > .05), for the aspiration type (u = 1602.0, p > .05), for the glottal type (u = 780.0, p > .05), and for the double type (u = 407.0, p > .05).

When NKs' overall performances on three stimulus types were compared, however, the percent correct for the aspiration type (80.33%) was considerably poorer than the glottal type (96.56%) and double type (95.00%). Friedman's tests showed that there was a significant difference among the three types ($\chi 2$[2df] = 15.750, p < .001), and Wilcoxon tests further revealed that their score for the aspiration type was significantly lower than the glottal type (z = −3.705, p < .001) and the double type (z = −3.252, p < .001), while there was no significant difference between their scores for the glottal and double types (z = −1.440, p > .05).

The results were further analyzed to examine the effects of the positiveness and the format within each stimulus type. Table 7 through 9 present the mean percentages of correct perception of the aspiration, glottal, and double types, respectively, for each format and positiveness.

Table 7. Mean Percentage of Correct NK Perception of Aspiration Type by Format and Positive/negativeness

Aspiration Format	Positive	Negative	Total Mean
VsC	95.00	83.33	89.17
CsC	83.33	66.67	75.00
CsCC	85.00	71.67	78.33
Total	87.78	73.89	80.83

Table 8. Mean Percentage of Correct NK Perception of Glottal Type by Format and Positive/negativeness

Glottal Format	Positive	Negative	Total Mean
VNV	100.00	96.25	98.13
VOV	98.75	91.25	95.00
Total	99.38	93.75	96.56

Table 9. Mean Percentage of Correct NK Perception of Double Type by Format and Positive/negativeness

Double Format	Positive	Negative	Total Mean
VLV	100.00	91.67	95.84
VCV	94.16		94.16
Total			95.00

For the aspiration type, Friedman's test showed that there was no significant effect of format ($\chi 2$[2df] = 5.115, $p > .05$). However, the difference between NKs' response to positive stimuli (87.78%) and negative stimuli (73.89%) was found to be significant : Wilcoxon signed-ranks tests confirmed that NKs' identified the items with a positive aspiration cue significantly better than the ones without it ($z = -2.632$, $p < .01$). For the glottal type, no significant difference was found between the two formats ($z = -1.890$, $p > .05$), but there was also an effect of positiveness ($z = -2.310$, $p < 0.05$). Similarly to the aspiration type, NKs performed significantly better on the items with a positive cue (99.38%) than on the ones without it (93.75%). Regarding the perception of the double type, no significant difference was found between the two formats ($z = -.577$, $p > .05$) nor between the positive and negative stimulus items in the format VLV ($z = -1.890$, $p > .05$).

To sum, Korean EFL learners generally performed very well in their perception task, irrespective of their English proficiency level. However, their performance was not consistent across all the stimulus types and positiveness. They performed significantly

worse in the perception of aspiration type than the other two types. Also, they did better in perceiving the items when the perceptual cues were present in the aspiration and glottal type stimuli.

4.3 Correlation between NKs' Production and Perception

In order to see if there is a significant correlation between NKs' production and perception scores, correlation analyses were made. Because normality of distribution could not be assumed, nonparametric Spearman rank-order correlations were calculated.

Table 10. Correlation between NKs' Production and Perception (N=20)

	Correlation Coefficient (Spearman's *Rho*)	*p*
Total scores	.341	.141
Aspiration scores	-.154	.518
Glottal scores	.315	.176
Double scores	.123	.606

There was no significant correlation between the NKs' production and perception scores in total (*rho* = .341, $p > .05$). The correlation was also examined for each of learners' aspiration, glottal, and double type scores, and no significant correlation was found for any of the three types between the production and perception. These results mean that, at least for the Korean EFL learners who participated in our experiments (at intermediate and advanced levels), the production and perception abilities of the target phrases involving word segmentation are not significantly correlated.

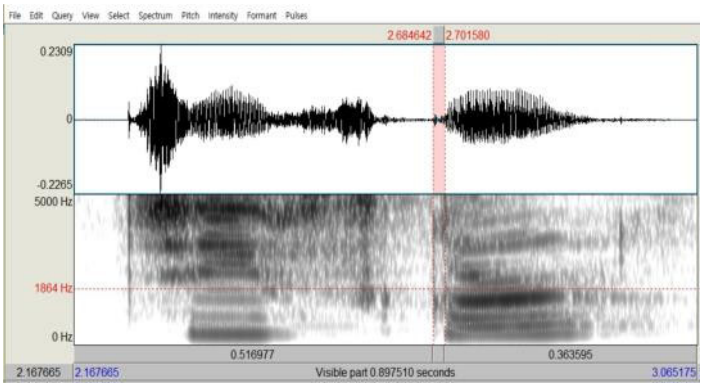
5. Discussion and Conclusion

To summarize the findings from the experiments conducted in this study, our production and perception experiments showed the followings: 1) our Korean subjects had more difficulty in the production of aspiration type than the other types; 2) they had more difficulty pronouncing the items with positive cues both in the aspiration type and glottal type stimuli; 3) compared to the results of the previous studies, our subjects generally performed well in their perception task, but they still did

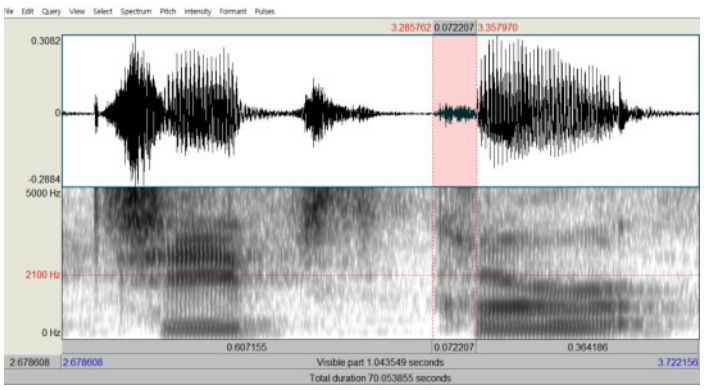
significantly worse in the perception of the aspiration type than the other two types; 4) unlike the results of the production task, they did better in perceiving the items when the positive perceptual cues were present both in the aspiration and glottal type stimuli; 5) the subjects' English proficiency level was not significantly associated with their performance in both production and perception; and 6) the production and perception abilities are not significantly correlated with each other.

The perception experiment was conducted mainly to see if there is a correlation between production and perception. Since the overall results (the findings 3 and 4 above) replicate those from the previous perception studies (Um 2006, Yun 2009, Chang and Park 2011) except that the subjects' general performances have improved, the discussion in this section will be focused on our subjects' production results.

Considering that the aspiration feature is contrastively used in Korean, it is interesting to note that our subjects were worst in producing the aspiration type stimuli (59.72% accuracy rate with average of 74.98% accuracy rate in production). In particular, they had more difficulty pronouncing positive aspiration stimuli than the negative (47.58% vs. 71.85% of accuracy rate respectively). This result may suggest that our Korean subjects had difficulty in producing word-initial stops after /s/ correctly with full aspiration, and thus, the English native speakers could not perceive them as aspirated, ending up with choosing the non-aspiration alternative from the answer choices (e.g, choosing *chief school* for *chief's cool*). Given that the length of VOT (Voice onset time: the duration of time interval between the articulatory release of the stop and the onset of vocal fold vibration) is thought to serve as a major cue in distinguishing aspirated and unaspirated, it can be hypothesized that our subjects did not produce intended word-initial voiceless stops with long enough VOTs. In fact, in her study on the comparison of native English speaker's voiceless stops and those of Korean speakers, Jo (2016) found that the mean VOT of voiceless stops for Korean speakers was shorter than for English speakers. According to Lisker and Abramson (1964), VOT of word-initial voiceless stops in English speakers ranges from 78.2 to 102.8 ms. (recited from Yun 2009). Let us take some examples of our subjects' production. The following shows the waveforms and spectrograms of *chief's cool* produced by our subjects.



(a) *chief's cool* produced by subject A



(b) *chief's cool* produced by subject B

Figure 1. The Waveforms and Spectrograms of *chief school*

Figure 1(a) is the waveform and spectrogram of *chief's cool* produced by one of our subjects whose score on positive aspiration items was the lowest. The VOT of the stop /k/ produced by the subject A is about 18 ms. It is too short a VOT to be perceived as aspirated in general. Consequently all three of our NE judges didn't perceive it correctly as *chief's cool* but perceived it as *chief school*. Figure 1(b) is the waveform and spectrogram of *chief's cool* produced by a subject whose score for aspirated stop production is the highest (70% correct rate). The VOT of the stop /k/ produced by the subject B is approximately 72 ms. and all three judges perceived the phrase correctly.

While contrasting the VOTs of the two subjects who were best and worst in the production of positive aspiration stimuli could provide a good example for the

correlation between the duration of VOT and NEs' perception of aspiration type stimuli, it is still not clear whether there is one to one correspondence between correctly produced items and the long length of VOT. Since the main goal of this study was to find out whether the native speakers understood the production of our subjects' phrases involving word boundaries as intended, the measuring and averaging of VOTs in all the subjects' target stops and comparing it with NEs' perception results were not done in this study, but I will rather leave the work to future studies.

Another related question that might be asked here is which range of VOT length can play a role as a perceptional cue to NEs. In case of the subject B who performed the best on the positive aspiration stimuli, his VOTs for word-initial voiceless stops (positive aspiration stimuli) range from 20 to 72 ms. (average 46.6 ms.), and his VOT for unaspirated target stops (negative aspiration stimuli) range from 0 to 50 ms. (average 25.89 ms.). Even the subject who performed well seems to have much shorter VOTs in word-initial voiceless stops compared to those of English speakers' (from 78.2 to 102.8 ms., according to Lisker and Abramson 1964). In the future studies, along with the measurement of VOTs, it would be interesting to further examine the relationship between the VOT ranges and NEs' perceptibilty of L2 learners' aspiration production and also other possible relevant phonetic details such as durations of the preceding segment and /s/ that might have contributed to NEs' perception.[7]

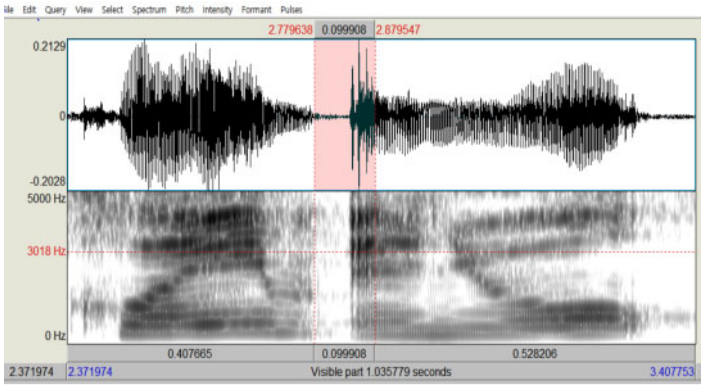
In terms of the L1 influence on L2 pronunciation, it is intriguing that Korean EFL learners have difficulty in producing aspiration for word-initial voiceless stops strongly enough to be correctly perceived by native speakers of English. Korean has aspirated consonants, of which VOT is quite long (91~126 ms. according to Lisker and Abramson 1964, recited from Yun 2009), in the phonemic inventory. Korean subjects' poor performance in the production of the positive aspiration stimuli, therefore, is not in accordance with our expectation. It might be explained in terms of the influence of L1 phonotactics/syllable structure and from the viewpoint of articulation. In Korean,

[7] Although aspiration has been recognized as a primary cue in English speech segmentation and Nakatani and Dukes (1977) found quantitative cues such as duration and amplitude did not provide strong juncture cues in English, Ito and Strange (2009), in their acoustic analysis of the stimuli recorded by an English native speaker, reported that the /s/ frication durations were noticeably shorter in the positive stimuli than in the negative stimuli. It would be interesting to see if the difference in the durations of /s/ frication or neighboring segments plays a role in signaling the word boundary both for native English speakers and for EFL learners. This task is left open for future research.

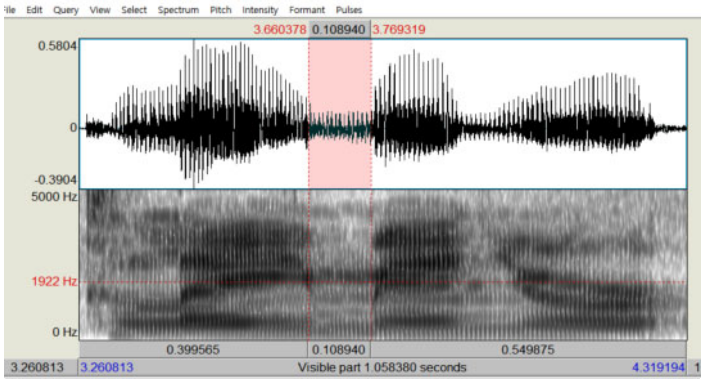
maximum number of consonants allowed in coda position is one. Therefore, if two underlying consonants occur in syllable-final position, one consonant is deleted. (e.g., /kaps/ 'price' is realized as [kap]). Also syllable-final /s/ is neutralized to [t] (e.g., /kos/ 'place' is realized as [kot]). Therefore, there is no /s/+stop clusters in Korean. So Koreans are not accustomed to producing fricative and stop consonant consecutively. English [s] is a high-pitched sound with an obvious hiss, and producing an aspirated voiceless stop after it (as in *keeps parking*) may involve considerable articulatory effort. For this reason, Korean subjects may have produced aspiration for word-initial voiceless stops following /s/ not as strongly as it has to be, with shorter VOTs. This might cause native English speakers to perceive their stops as unaspirated, misperceiving, for example, *keeps parking* as *keep sparking*.

With regard to the production results of the glottal type stimulus, native English speakers perceived what our subjects said quite well, better than the aspiration type, but they also had more difficulty perceiving the items that necessitate positive glottal cue (positive stimuli) than the ones that don't need it (negative stimuli) (74.17% vs. 87.08% respectively). Given that glottal stop/glottalization is thought to be universally easy to perceive, this result may suggest that our subjects' production of these items was not fully accompanied by glottal stop/glottalization which is strong enough to be a phonetic cue to native speakers.

Acoustically, glottal stop is identified by a drop in amplitude, abruptly changed periodicity and disruption of vocalization (Olive et al. 1993). Figure 2(a) below shows the waveform and spectrogram of *claim annual* correctly produced by one of our subjects. As can be seen from the shaded part, there is a gap before the beginning of the vowel and also abrupt change of periodicity in the first part of the vowel. All three judges correctly perceived it. Figure 2(b) is the waveform and spectrogram of *claim annual* produced by another subject. As can be seen in the shaded part, this subject seemed to produce this phrase without inserting glottal stop or a gap between the two words. All three NE judges did not perceive it as *claim annual,* but perceived it as *clay manual*.



(a) *claim annual* produced by subject C



(b) *claim annual* produced by subject D

Figure 2. The Waveforms and Spectrograms of c*laim annual*

This discrepancy of accuracy in the NEs' perception between positive and negative glottal items produced by our subjects may also be explained by L1 transfer. Unlike English where glottal stop is often inserted before word-initial vowels, in Korean, onsetless syllables are allowed and glottal stop is not usually inserted before syllable-initial vowel.[8] Therefore, influenced by their L1, some Korean learners tend not to insert glottal stop before a word-initial vowel, (for example, linking *seen* and *either* in *seen either* without inserted glottal stop between the two words).

[8] Though glottal stop insertion before a word-initial vowel may not be an obligatory phonological rule in English, the native English speaker who recorded target items in our study pronounced the target items consistently with an inserted glottal stop.

Speaking shortly of the perception experiment with Korean learners, it was found that, compared to the results of the previous perception studies with Korean EFL learners, the subjects' overall performance on the present word boundary segmentation task has highly improved with around 90% overall accuracy rate (80.8% accuracy rate on aspiration type stimuli). Of course direct comparison should be made with caution because the experiments were conducted with different subjects and in different conditions. And yet it is noteworthy that the subjects performed well irrespective of their English proficiency level. In the previous perception study by Um (2006), it was shown that the intermediate group performed significantly worse than the advanced group. In particular, the difference was found in their perception of the negative target items. Such a discrepancy was not observed in the present study, which suggests that the ability of intermediate group in utilizing acoustic phonetic cues, especially negative cues, has also considerably improved. One of the reasons might be the fact that Korean learners are exposed to different English learning environments from the past when the previous research was conducted. They tend to be exposed to English materials presented with audio-visual multimedia more than before and at an earlier age. This might contribute to the learners' acquisition of phonetic cues at the earlier stage of English learning. Of course, this is a simple conjecture about one of the reasons, and there might be other reason(s) to which the present results of the perception task are attributed.[9]

As for the question of correlation between production and perception, the present study did not find a correlation of Korean EFL learners' production and perception abilities at least in distinguishing word boundaries using phonetic variations and acoustic phonetic cues. The result seems to tell us that among L2 English learners who have reached a very high level of perceptive ability, a difference in perception ability does not necessary cause a difference in production ability. The present study, however, could not answer the question of whether L2 learners with lower perception ability will show the correlation between in their perception and production abilities.

Lastly, it is noteworthy that while our subjects performed very well in perceiving positive stimuli at about 95% correct rate on average, they could correctly produce the

[9] Because the primary goal of this study was to investigate NK's production of the target language, participants were given the production task before the perception task to eliminate any potential influence of the prior perception task. However, as pointed out by an anonymous reviewer, there also exists the possibility that the better performance of the subjects' perception test was partly due to familiarization with the given stimuli from their perception task.

same positive stimuli at about 70% rate on average. These results suggest that many subjects who were able to utilize acoustic phonetic cues to perceive word boundaries have failed to use the same word boundary cues for their production of target phrases. These results are consistent with the 'perception before production' view in L2 phonology which has received support from a number of empirical studies (e.g., Flege 1991, 1993). In this sense, while the current study did not present a positive correlational mapping between perception and production, it still contributes to finding out the ordering relation of perception and production in the developmental course of L2 phonology acquisition.

References

Altenberg, E. 2005. The perception of word boundaries in a second language. *Second Language Research* 24(4), 325-358.

Chang, S and H. Park. 2011. A study of the use of allophonic cues in the perception of English word boundaries by Korean learners of English. *Phonetics and Speech Sciences* 3(3), 63-68.

Christie, W. 1974. Some cues for syllable juncture perception in English. *Journal of the Acoustical Society of America* 55, 819-21.

Flege, J. E. 1991. Perception and production: The relevance of phonetic input to L2 phonological learning. In T. Hueber and C. Ferguson, eds., *Crosscurrents in Second Language Acquisition and Linguistic Theories,* 249-289. Amsterdam: John Benjamins.

Flege, J. E. 1993. Production and perception of a novel, second-language phonetic contrast. *Journal of Acoustical Society of America* 93(3), 1589-1608.

Ito, K. and W. Strange. 2009. Perception of allophonic cues to English word boundaries by Japanese second language learners of English. *Journal of Acoustical Society of America* 125(4), 2348-2360.

Jo, J. 2016. Native and non-native English speakers' production of voiceless stops: effects of lexical stress and prosodic domain on VOT. *English Studies* 36, 111-128.

Kent, R. and C. Read. 1992. *The Acoustic Analysis of Speech*. San Diego: Singular Publishing Group.

Nakatani, L. and K. Dukes. 1977. Locus of segmental cues for word juncture. *Journal*

of Acoustical Society of America 62, 714-719.
Lisker, L. and A. Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20, 384-422.
Olive, J., A. Greenwood and J. Coleman. 1993. *The Acoustics of American English Speech: A Dynamic Approach*, NY: Springer-Verlag.
Schwartz, G. 2016. Word boundaries in L2 speech: Evidence from Polish learners of English. *Second Language Research* 32(3), 397-426.
Shin, S. and Y. Hwang. 2012. Perception and production of English geminate consonants across word boundaries by Korean learners and native speakers of English. *Studies in Phonetics, Phonology and Morphology* 18(1), 85-110.
Shoemaker, E. 2014. The exploitation of subphonemic acoustic detail in L2 speech segmentation. *Studies in Second Language Acquisition* 36, 709-731.
Yun, G. 2009. Korean learners' perceptual cues for word boundaries in English. *Korean Journal of Linguistics* 34(1), 73-92.
Um, H. 2006. The perception of word boundaries by Korean college EFL learners. *The Linguistic Association of Korea Journal* 14(3), 51-70. .

Examples in: English
Applicable Languages: English
Applicable Level: Tertiary

Um, Hye-Young, Professor
Myongji University
Bangmok College of General Education
116 Myongji-ro, Cheoin-gu, Yongin-si
Gyeonggi-do, 17058, Korea
Tel: 031-330-6264
E-mail: hyum@mju.ac.kr

Received: February 7, 2020
Revised: March 10, 2020
Accepted: March 20, 2020