pISSN 2005-8063
eISSN 2586-5854
2020. 3. 31.
Vol.12 No.1
pp. 43-50

말소리와 음성과학

Phonetics and Speech Sciences

한국음성학회지

https://doi.org/10.13064/KSSS.2020.12.1.043

Check for updates

Effects of base token for stimuli manipulation on the perception of Korean stops among native and non-native listeners*

Eunjin Oh**

Department of English Language and Literature, Ewha Womans University, Seoul, Korea

Abstract

This study investigated whether listeners' perceptual patterns varied according to base token selected for stimuli manipulation. Voice onset time (VOT) and fundamental frequency (F0) values were orthogonally manipulated, each in seven steps, using naturally produced words that contained a lenis (/kan/) and an aspirated (/k^han/) stop in Seoul Korean. Both native and non-native groups showed significantly higher numbers of aspirated responses for the stimuli constructed with /k^han/, evidencing the use of minor cues left in the stimuli after manipulation. For the native group the use of the VOT and F0 cues in the stop categorization did not differ depending on whether the base token included the lenis or aspirated stop, indicating that the results of previous studies remain tenable that investigated the relative importance of the acoustic cues in the native listener perception of the Korean stop contrasts by using one base token for manipulating perceptual stimuli. For the non-native group, the use patterns of the F0 cue differed as a function of base token selected. Some findings indicated that listeners used alternative cues to identify the stop contrast when major cues sound ambiguous. The use of the manipulated VOT and F0 cues by the non-native group was not native-like, suggesting that non-native listeners may have perceived the minor cues as stable in the context of the manipulated cue combinations.

Keywords: base token, stimuli manipulation, perception, Korean stops

1. Introduction

A phonetic contrast in a language is in general signaled by multiple acoustic cues, and the relative importance of the cues differs in the perception of the phonetic contrast (e.g., Abramson & Lisker, 1985). To estimate the relative importance of the cues, the acoustic cues under investigation can be systematically synthesized, and the degrees to which cues contribute to the identification of the phonetic contrast can be quantified. Shultz et al. (2012) examined the relative weight of voice onset time (VOT) and fundamental frequency (F0) cues in the perception of English stop voicing. They constructed perceptual stimuli by orthogonally synthesizing the VOT (ten steps) and the onset F0 (four steps) values and controlling other cues (e.g., burst duration, aspiration amplitude, vowel length, and formant transition) as being linearly correlated with the VOT values. Listeners weighted the VOT cue more heavily than the F0 cue when categorizing the English stop voicing.

* This research was supported by the Intramural Research Grant of Ewha Womans University.

** ejoh@ewha.ac.kr, Corresponding author

Received 15 February 2020; Revised 11 March 2020; Accepted 11 March 2020

ⓒ Copyright 2020 Korean Society of Speech Sciences. This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Another method for designing perceptual stimuli is to manipulate relevant cues based on naturally produced tokens. Here, all cues except those under investigation remain in the stimuli, which will thereby likely sound more natural than the fully synthesized stimuli described above. McGuire (2010: 11) referred to this kind of stimuli as "naturalistic but controlled stimuli." A potential problem with this method is that the cues left in the stimuli may affect perception. Above all, they may cause a response bias in which the number of responses is inclined toward the base token used for stimuli manipulation. That is, in the perception of a phonetic contrast A and B, when the stimuli are manipulated with a base token containing A, the number of A responses is likely to be higher. In addition, mismatches between manipulated cues and those left in the stimuli may also affect the identification of the phonetic contrast. To investigate the patterns of using VOT and F0 cues in the perception of stop contrasts in English and Korean by English learners of Korean, Kong & Edwards (2015) constructed English stimuli by cutting the VOT part from a /ta/ token and attaching it in six steps in front of the voicing onset of a /da/ token. The mismatches between the aspiration information associated with the voiceless stop and the information in the following vowel associated with the voiced stop (e.g., onset F0, H1–H2, F0 contour shape, and vowel duration) could have affected the use patterns of the perceptual cues in some manner.

The purpose of the present study was to investigate whether the cues remaining intact in the stimuli manipulated from a natural token affect the listener perception of a phonetic contrast. Specifically, this study explored (1) whether the cues left in the stimuli caused a response bias toward the base token and (2) whether the use patterns of the manipulated cues varied according to the base token selected. The case study examined the use of VOT and F0 cues in the perception of lenis and aspirated stops in Seoul Korean.

Seoul Korean has three phonation types in its stop system, fortis *vs*. lenis/aspirated stops differentiated mainly as short-lag *vs*. long-lag VOTs, and lenis *vs*. fortis/aspirated stops as low *vs*. high F0s. The VOT cue is losing its contrasting role between the lenis and aspirated stops and F0 is instead functioning as the most important cue for the stop contrast (e.g., Kang, 2014; Kim et al., 2002; Silva, 2006). Oh et al. (2018) investigated whether VOT is a relatively less important cue for females than males in the perception of the lenis and aspirated stops. This question was based on the results of previous studies demonstrating that female speakers showed significantly smaller VOT distinctions between the two stops than male speakers (e.g., Kang, 2014; Oh, 2011). In the perception study, the researchers constructed stimuli by manipulating the VOT and F0 values with a word containing an aspirated stop (/k^{h}an/). Therefore, the stimuli may have contained stop cues other than VOT and F0 that could potentially bias responses toward the aspirated stop (also see Lee et al., 2013, which used a word that contained a lenis stop as the base token for stimuli manipulation to explore the relative weightings of VOT and F0 in the perception of Korean stops as a function of dialects).

In the present study, the VOT and F0 values were manipulated with monosyllabic words that contained a lenis (/kan/) and an aspirated stop (/k^{h}an/), respectively, to examine whether perceptual patterns vary as a function of the base token. Acoustic cues to the stop contrast other than VOT and F0 (e.g., closure duration, burst amplitude, aspiration amplitude, H1–H2, F0 values in the following vowel, and vowel duration) were left unchanged in the stimuli. Whether the numbers of responses vary and whether the usage patterns of the VOT and F0 cues differ according to the base tokens were examined. Presented with the orthogonally varied VOT and F0 values, it was possible that listeners paid closer attention to the variation of the two major cues and thus became desensitized to other minor cues left in the stimuli (cf., Winn et al., 2012).

This study conducted the same perception experiment for non-native listeners of Korean and examined whether the sensitivity to the minor stop cues left in the stimuli differed between native and non-native listeners. Based on the assumption that native listeners would have more knowledge of the minor cues that signal the phonetic contrast than non-native listeners, it was hypothesized that native listeners' responses would be more inclined toward the selected base token than those of non-native listeners.

2. Method

2.1. Subjects

Forty subjects were paid for their participation in the perception experiment. Twenty were native speakers of Seoul Korean (mean age 22.0 years; age range 20–27 years) and 20 were non-native speakers of Korean whose native language was Mandarin Chinese (mean age 24.7 years; age range 19–33 years). All were female and, at the time of the experiment, either undergraduate or graduate students at a university in Seoul. No participants reported any speech or hearing disorders. The native listeners were born and raised in Seoul and had no experience of residing abroad for more than a year. The non-native listeners were born and raised in mainland China and had no experience of residing abroad other than in Korea for more than six months. Their TOPIK (Test of Proficiency in Korean; administered by the National Institute for International Education) levels were between 4 and 6 (mean 5.4), with level 6 corresponding to the highest linguistic competence.

2.2. Base Tokens and Stimuli Manipulation

The base tokens were selected from the production data pool in Oh et al. (2018). The tokens were /kan/ and /k^{h}an/, produced by a female native speaker of Seoul Korean who was born and raised in Seoul (age 22 years; an undergraduate student at a university in Seoul at the time of recording; experience residing abroad for five months). The two monosyllabic words were produced in a carrier sentence of "tasi 'kan/k^{h}an' poseyo (see 'kan/k^{h}an' again)." The base tokens were excised from after /i/ of /tasi/ to 25 ms after the coda nasal. As for the acoustic values of the tokens, VOT was 75 ms for /kan/ and 76 ms for /k^{h}an/; onset and mid F0 in the following vowel were 220 Hz and 208 Hz, respectively, for /kan/, and 296 Hz and 296 Hz, respectively, for /k^{h}an/; the closure duration measured from after /i/ of /tasi/ to the release burst was 56 ms for /kan/ and 89 ms for /k^{h}an/; and [vowel+nasal] duration was 234 ms for /kan/ and 198 ms for /k^{h}an/.

Stimuli were synthesized using the PSOLA function in Praat (Boersma & Weenink, 2016). The VOT and F0 values were each manipulated in seven equal steps (VOTs 30–90 ms in 10 ms steps and F0s 180–312 Hz in 22 Hz steps) for both the stimuli constructed with /kan/ (/kan/-based stimuli, hereafter) and those

constructed with /k^{h}an/ (/k^{h}an/-based stimuli, hereafter). These VOT and F0 ranges reflected the minimum and maximum values found in the female data pool in Oh et al. (2018). The VOT durations were either contracted or expanded to the intended values using the relative duration function, and the F0 values at the vowel midpoint were lowered or raised to the intended values. The F0 contour shapes in the base tokens were maintained, potentially functioning as an additional biasing cue. For each base token, the seven VOT steps were created first and then the seven F0 steps were created for each VOT level, resulting in 98 stimuli in total (7 VOTs×7 F0s×2 base tokens).

2.3. Procedure

The perception experiment was conducted using the MFC function in Praat (March and April, 2019). The 98 stimuli with five repetitions constituted 490 trials in total. The /kan/- and /k^{h}an/-based stimuli were blocked separately, and the presentation order of the two blocks was counterbalanced between listeners. The stimuli were randomized for each listener. A randomization method was used to arrange the same stimuli non-consecutively. Listeners listened through Audio-Technica ATH-M40X headphones in a quiet office and completed a forced-choice identification task by clicking on either /kan/ or /k^{h}an/ in Korean orthography on a computer screen. Listeners had no time limit for choosing the answer, but could not modify their answers after selection. They completed a short practice session with eight stimuli selected from the overall stimuli pool. Each listener took approximately 20–25 minutes to complete the task.

2.4. Statistical Analyses

To model the effects of the stop cues (VOT or F0) and the base tokens, mixed-effects binary logistic regression analyses were performed separately for the data from native and non-native participants using the program R (R Development Core Team, 2016) and the package "lme4" (Bates et al., 2015). The VOT and F0 values were coded as levels (−3, −2, −1, 0, 1, 2, and 3). The dependent variable was Response. VOT, F0, and Base Token were the fixed effects, and Subject was the random effect. The interaction terms VOT×Base Token and F0×Base Token were included to investigate whether the use patterns of the VOT and F0 cues differed depending on base token selected for stimuli manipulation [Response~VOT+F0+Base Token+VOT×Base Token+F0×Base Token+(1|Subject)]. The reference for the base token was /k^{h}an/ (section 3.1).

To investigate whether there were differences between the native and non-native group in the use of the manipulated VOT and F0 cues, the data of the two groups were merged, and mixed-effects binary logistic regression analyses were performed with the merged data separately for the /kan/- and the /k^{h}an/-based stimuli (section 3.2). The dependent variable was Response. VOT, F0, and Listener Group were the fixed effects and Subject was the random effect. The interaction terms VOT×Listener Group and F0×Listener Group were included to investigate whether the use of the VOT and F0 cues differed between the native and non-native groups [Response~VOT+F0+Listener Group+VOT×Listener Group+F0×Listener Group+(1|Subject)]. The reference for the listener group was the native group.

3. Results

3.1. Effects of Base Token on Stop Categorization

Figure 1 displays the percentages identified as /k^{h}an/ for the VOT (left) and F0 steps (right). The top panels show native data and the bottom panels show non-native data. The circles and triangles represent the responses to the /kan/- and /k^{h}an/-based stimuli, respectively.

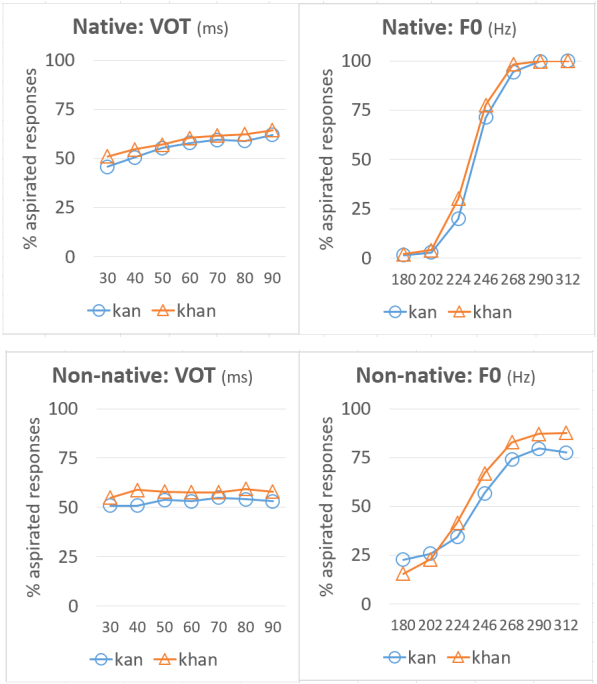

Figure 1. Percentages identified as /k^{h}an/ for the VOT (left) and F0 steps (right) to /kan/- (circles) and /k^{h}an/-based stimuli (triangles) by native (top) and non-native listeners (bottom).

For the native group, as the VOT and F0 values increased, the percentages of the /k^{h}an/ responses increased. The outputs of the mixed effects logistic regression models are summarized in Table 1. Both VOT (β=0.44635, *SE*=0.03347, z=13.335, p<0.0001) and F0 (β=2.56487, *SE*=0.08233, z=31.152, p<0.0001) were significant predictors of the response patterns. The estimated coefficient value of F0 was larger than that of VOT, which indicates that F0 is the more important cue in categorizing the lenis and aspirated stops, in agreement with previous research results (e.g., Kang, 2010; Kim et al., 2002; Lee et al., 2013; Oh et al., 2018).

The percentages of aspirated responses to the /k^{h}an/-based stimuli were higher than those to the /kan/-based stimuli for all VOT and F0 steps except the 6th and 7th steps of F0, for which the aspirated response rates were the same to the /k^{h}an/- and /kan/-based stimuli as 99.7% and 100%, respectively. This effect of the base token was statistically significant (β=0.54422, *SE*=0.10259, z=5.305, p<0.0001). The coefficient value for the base token was positive, indicating that the /k^{h}an/-based stimuli elicited significantly more aspirated responses than the /kan/-based stimuli. This result suggests that the stop cues other than VOT and F0 played a role in the listeners' categorization of the two stops.

The aspirated responses to the /k^{h}an/-based stimuli were more than 50% at all steps of VOTs. In the case of the /kan/-based stimuli, the

perceptual crossover point between /k/ and /k^h/ for VOT was at around 40 ms (the top left panel of Figure 1). For F0, the perceptual crossover point shifted slightly leftward for the /k^han/-based stimuli, indicating that the more aspirated responses started with a smaller F0 value for the /k^han/-based stimuli compared to the /kan/-based stimuli (the top right panel of Figure 1).

Table 1. Parameter statistics for base token models for the native (top) and non-native group (bottom)

Native group	Estimate	Std. error	z value	Pr (>|z|)
(Intercept)	1.01916	0.20187	5.049	<0.0001
VOT	0.44635	0.03347	13.335	<0.0001
F0	2.56487	0.08233	31.152	<0.0001
Base token	0.54422	0.10259	5.305	<0.0001
VOT×Base token	–0.06622	0.04628	–1.431	0.152
F0×Base token	–0.02131	0.10983	–0.194	0.846
Non-native group	Estimate	Std. error	z value	Pr (>|z|)
(Intercept)	0.15201	0.09069	1.676	0.0937
VOT	0.02690	0.01625	1.656	0.0978
F0	0.52874	0.01792	29.508	<0.0001
Base token	0.32272	0.04882	6.611	<0.0001
VOT×Base token	–0.00517	0.02406	–0.215	0.8298
F0×Base token	0.21559	0.02772	7.777	<0.0001

Neither the interaction between VOT and Base Token (β=–0.06622, SE=0.04628, z=–1.431, p=0.152) nor that between F0 and Base Token (β=–0.02131, SE=0.10983, z=–0.194, p=0.846) was statistically significant. The differences between the maximum and minimum percentages of the aspirated responses according to the VOT steps were 16.3% (62.0% at the 7th step – 45.7% at the 1st step) and 13.3% (64.3% at the 7th step – 51.0% at the 1st step) for the /kan/- and /k^han/-based stimuli, respectively. Those differences according to the F0 steps were 98.6% (100% at the 7th step – 1.4% at the 1st step) and 98.1% (100% at the 7th step – 1.9% at the 1st step), respectively. These results indicate that the use patterns of the VOT and F0 cues in categorizing the lenis and aspirated stops did not differ by whether /kan/ or /k^han/ was the base token used for stimuli manipulation.

In the case of the non-native group, as the VOT values increased the percentages of the /k^han/ responses did not increase, and the effect of VOT was not statistically significant (β=0.02690, SE=0.01625, z=1.656, p=0.0978). As the F0 values increased the percentages of the /k^han/ responses increased, and the effect of F0 was statistically significant (β=0.52874, SE=0.01792, z=29.508, p<0.0001). That is, the non-native group used the F0 cue, but not the VOT cue, to distinguish between the lenis and aspirated stops in Seoul Korean. Non-native listeners were able to use the F0 cue, but not to the same extent as native listeners. The estimate coefficient value was smaller for the non-native group (0.52874) compared to the native group (2.56487).

The percentages of the aspirated responses to the /k^han/-based stimuli were higher than those to the /kan/-based stimuli for all VOT and F0 steps except the 1st and 2nd steps of F0, for which the aspirated response rates were higher for the /kan/-based stimuli than the /k^han/-based stimuli. This effect of the base token was statistically significant (β=0.32272, SE=0.04882, z=6.611, p< 0.0001). The coefficient value for the base token was positive, indicating that the /k^han/-based stimuli elicited significantly more aspirated responses than the /kan/-based stimuli. This result indicates that the stop cues other than VOT and F0 were involved in the stop categorization also for the non-native group.

The aspirated responses were more than 50% at all steps of VOTs for both /kan/- and /k^han/-based stimuli (the bottom left panel of Figure 1). For F0, the perceptual crossover point between /k/ and /k^h/ shifted leftward for the /k^han/-based stimuli, indicating that more aspirated responses started with a smaller F0 value for the /k^han/-based stimuli compared to the /kan/-based stimuli also for the non-native group (the bottom right panel of Figure 1).

The interaction between VOT and Base Token was not significant (β=–0.00517, SE=0.02406, z=–0.215, p=0.8298), which indicates that the use of the VOT cue did not differ by whether the base token was /kan/ or /k^han/. The differences between the maximum and minimum percentages of the aspirated responses according to the VOT steps were 4.0% (54.9% at the 5th step – 50.9% at the 1st and 2nd steps) and 4.5% (59.4% at the 6th step – 54.9% at the 1st step) for the /kan/- and /k^han/-based stimuli, respectively. The interaction between F0 and Base Token was statistically significant (β=0.21559, SE=0.02772, z=7.777, p<0.0001). The coefficient value was positive, indicating that the non-native group used the F0 cue more efficiently when the base token was /k^han/. The differences between the maximum and minimum percentages of the aspirated responses according to the F0 steps were 56.9% (79.6% at the 6th step – 22.7% at the 1st step) and 72.3% (87.7% at the 7th step – 15.4% at the 1st step) for the /kan/- and /k^han/-based stimuli, respectively.

In sum, the differences between the native and non-native groups were (1) whether the VOT cue was used to distinguish between the lenis and aspirated stops and (2) whether the use of the F0 cue differed depending on what base token was selected to manipulate the perceptual stimuli. The native group, but not the non-native group, used the VOT cue to categorize the two stops, and the use patterns of the F0 cue did not differ depending on whether the base token was /kan/ or /k^han/ for the native group but did differ for the non-native group.

Figure 2 shows the percentages of aspirated responses for each F0 step in order to examine the usage patterns of the VOT cue for each F0 step by the native (top) and non-native group (bottom). The left panels show data from the /kan/-based stimuli and the right panels show data from the /k^han/-based stimuli. For the native group, at the unambiguous F0 steps of 180 Hz and 202 Hz, the response rates with /k^han/ were generally close to 0% for all VOT steps. At the unambiguous F0 steps of 268 Hz, 290 Hz, and 312 Hz, the response rates with /k^han/ were generally close to 100% for all VOT steps. On the other hand, at the ambiguous F0 steps of 224 Hz and 246 Hz, the native group used the VOT cue more systematically, which indicates that native listeners used VOT as an alternative cue when the F0 cue was ambiguous. The percentages of the /k^han/ responses increased overall from the 1st to the 7th VOT steps: 5%, 8%, 19%, 23%, 24%, 24%, and 36% at 224 Hz and 32%, 57%, 73%, 82%, 84%, 84%, and 88% at 246 Hz for the /kan/-based stimuli (the top left panel), and 10%, 22%, 22%, 34%, 42%, 39%, and 41% at 224 Hz and 54%, 63%, 77%, 85%, 83%, 90%, and 91% at 246 Hz for the /k^han/-based stimuli (the top right panel).

The non-native group showed different patterns from the native group in using the VOT cue for each F0 step. At the unambiguous F0 steps of 180 Hz and 202 Hz, the response rates with /k^han/ were not close to 0%, and at the unambiguous F0 steps of 268

Hz, 290 Hz, and 312 Hz, the response rates with /k^{h}an/ were not close to 100% (the bottom two panels). In addition, at the ambiguous F0 values of 224 Hz and 246 Hz, the aspirated responses did not increase systematically as the VOT values increased: 36%, 29%, 33%, 34%, 42%, 31%, and 35% at 224 Hz and 49%, 57%, 58%, 57%, 58%, 59%, and 59% at 246 Hz for the /kan/-based stimuli (the bottom left panel), and 34%, 42%, 44%, 39%, 48%, 38%, and 46% at 224 Hz and 59%, 66%, 67%, 68%, 65%, 73%, and 70% at 246 Hz for the /k^{h}an/-based stimuli (the bottom right panel). These results indicate that native listeners used the VOT cue as an alternative cue when the F0 cue was ambiguous and this phenomenon did not happen to non-native listeners (see section 4 for discussion).

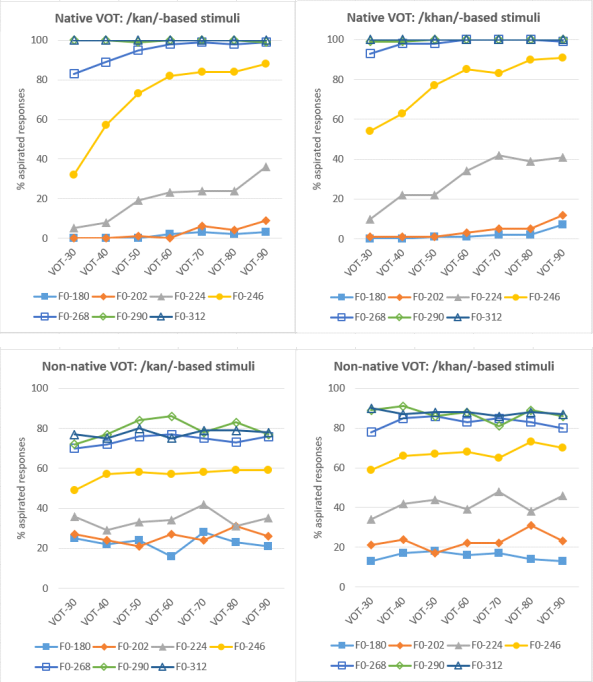

Figure 2. Percentages identified as /k^{h}an/ for each F0 step to /kan/- (left) and /k^{h}an/-based stimuli (right): Native (top) *vs*. non-native listeners (bottom).

3.2. Individual Listener Patterns Using Stop Cues

Figure 3 displays scatterplots for the numbers of the aspirated responses to the /kan/- (x-axis) and the /k^{h}an/-based stimuli (y-axis) by individual native (circles) and non-native (triangles) listeners. Most data points are located at the top left of the y=x line, indicating that most individual listeners responded with /k^{h}an/ more to the /k^{h}an/-based stimuli than to the /kan/-based stimuli. This result suggests that most of the native and non-native individual listeners used the minor stop cues left in the stimuli after the cue manipulation in differentiating the lenis and aspirated stops.

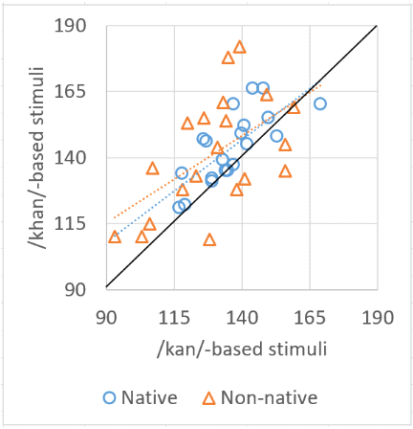

Figure 3. Scatterplots for the numbers of aspirated responses to the /kan/- (x-axis) and the /k^{h}an/-based stimuli (y-axis) by individual native (circles) and non-native listeners (triangles).

Figure 4 presents individual listener patterns using the manipulated VOT (left) and F0 cues (right) in the stop categorization as a function of the listener group. The above panels show the native data and the below panels show the non-native data. In the case of the native group, the cue-use patterns appear relatively consistent across individual listeners, whereas those of the non-native group showed a high amount of variability. Some non-native listeners showed similar patterns to native listeners, while others showed considerably different cue-use patterns. Especially for F0 of some non-native listeners, the lower the F0 value, the more they answered as aspirated (the bottom right panel in Figure 4).

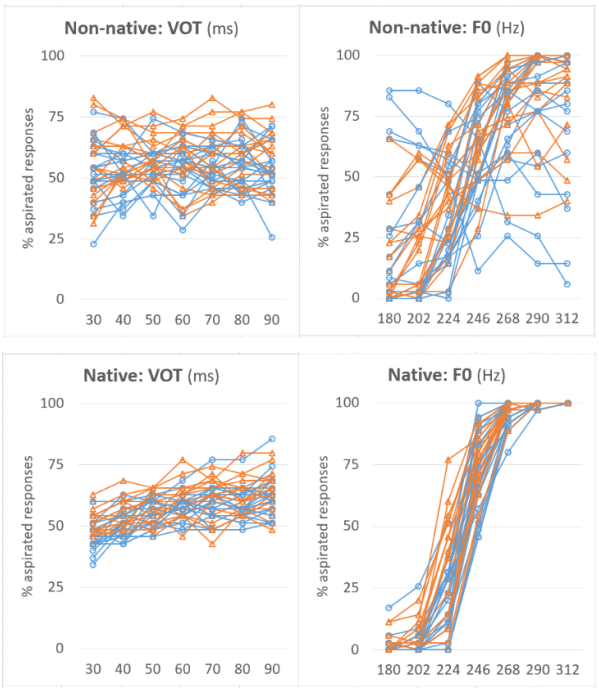

Figure 4. Individual listener patterns using the VOT (left) and F0 cues (right) in the stop categorization: native (top) *vs*. non-native data (bottom). Circles (blue) and triangles (orange) represent the percentages of the aspirated responses to the /kan/- and /k^{h}an/-based stimuli, respectively.

To examine whether there were differences between native and non-native listeners in the use of the manipulated VOT and F0 cues, mixed-effects binary logistic regression analyses were performed separately for the /kan/- and the /k^{h}an/-based stimuli (section 2.4). Table 2 presents the parameter statistics for the listener group models for the /kan/-based stimuli (above) and the /k^{h}an/-based stimuli (below). For the /kan/-based stimuli, both interactions between VOT and Listener Group (β=0.41574, SE=0.03733, z=11.138, p<0.0001) and between F0 and Listener Group (β=2.01395, SE=0.08748, z=23.021, p<0.0001) were statistically significant. Likewise for the /k^{h}an/-based stimuli, both interactions between VOT and Listener Group (β=0.35977, SE=0.03743, z=9.611, p<0.0001) and between F0 and Listener Group (β=1.79179, SE=0.08876, z=20.187, p<0.0001) were statistically significant. These results indicate that non-native listeners' use of the manipulated VOT and F0 cues in categorizing the lenis and aspirated stops were not native-like.

Table 2. Parameter statistics for listener group models for the /kan/-based stimuli (above) and the /khan/-based stimuli (below)

/kan/-based stimuli	Estimate	Std. error	z value	Pr (>|z|)
(Intercept)	0.15243	0.14514	1.050	0.2936
VOT	0.02692	0.01625	1.656	0.0977
F0	0.52910	0.01800	29.396	<0.0001
Listener group	0.85820	0.21415	4.007	<0.0001
VOT×Listener group	0.41574	0.03733	11.138	<0.0001
F0×Listener group	2.01395	0.08748	23.021	<0.0001
/k^{h}an/-based stimuli	Estimate	Std. error	z value	Pr (>|z|)
(Intercept)	0.48830	0.16689	2.926	<0.005
VOT	0.02231	0.01798	1.240	0.21487
F0	0.76517	0.02214	34.563	<0.0001
Listener group	1.08330	0.24650	4.395	<0.0001
VOT×Listener group	0.35977	0.03743	9.611	<0.0001
F0×Listener group	1.79179	0.08876	20.187	<0.0001

Figure 5 shows the identification patterns of some individual listeners displaying different cue-use patterns than native listeners. NNL-1 and NNL-15 were non-native listeners who used the F0 cue more effectively when /kan/ was the base token compared to when /k^{h}an/ was the base token. This contrasts with the results for the non-native group level, which used F0 more efficiently when /k^{h}an/ was the base token (section 3.1). Meanwhile, NNL-12 used the F0 cue more effectively when /k^{h}an/ was the base token than when /kan/ was the base token. For NNL-10, NNL-14, and NNL-19, the rates of aspirated responses increased as the F0 values increased when /k^{h}an/ was the base token, but the rates of the aspirated responses decreased as the F0 values increased when /kan/ was the base token. For NNL-13, the rates of the aspirated responses decreased as the F0 values increased when /kan/ was the base token and when /k^{h}an/ was the base token. NNL-11 and NNL-20 did not use the F0 cue in a native-like manner to distinguish between the two stops when either /kan/ or /k^{h}an/ was the base token.

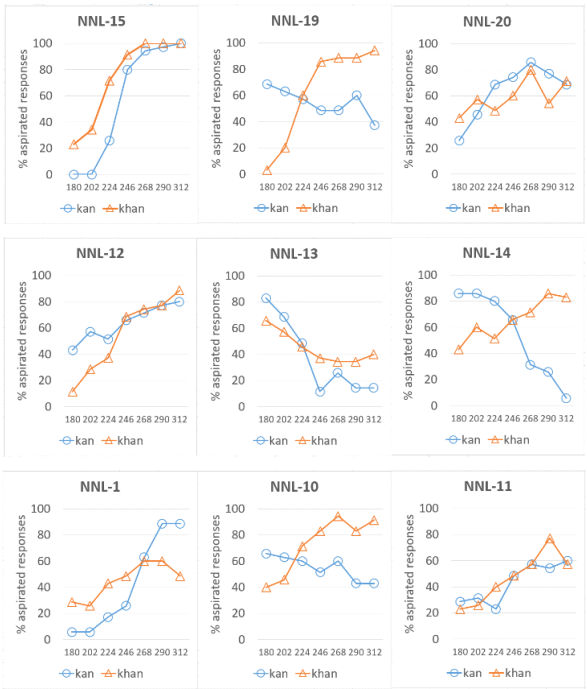

Figure 5. Identification patterns of some individual non-native listeners (NNLs) to the /kan/- (circles) and /k^{h}an/-based stimuli (triangles).

4. Summary and Discussion

The purpose of this study was to investigate whether listeners' perceptual patterns varied as a function of the base token selected for stimuli manipulation when categorizing the lenis and aspirated stops of Seoul Korean. The effect of base token selection was statistically significant for both native and non-native groups; the number of aspirated responses was significantly larger when the base token contained the aspirated stop. This result indicates that minor stop cues left in the stimuli after manipulating the VOT and F0 values were used in the stop categorization by both native and non-native listeners.

The differences between the total aspirated responses to the /k^{h}an/- and /kan/-based stimuli were 152 for the native group (2,880–2,728) and 236 for the non-native group (2,831–2,595). That is, the difference in the numbers of responses according to the base tokens was rather larger for the non-native group than the native group. The hypothesis that the native group would be more inclined in its responses toward the base tokens under the assumption that native listeners would have more knowledge of the minor cues signaling the stop contrast was not supported.

For the native group, the interactions between the acoustic cues and the base token were not statistically significant, indicating that the use of the VOT and F0 cues in the stop categorization did not differ depending on whether the base token included the lenis or aspirated stop. This finding suggests that the results of previous studies remain tenable that investigated the relative importance of the acoustic cues in the native listener perception of the Korean stop contrasts by using one base token for manipulating perceptual stimuli (e.g., Lee et al., 2013; Oh et al., 2018). For the non-native group, the interaction between F0 and the base token was statistically significant, indicating that the use patterns of the F0 cue differed as a function of base token selected. The F0 cue was

used more efficiently when the base token was /k^han/. This finding may suggest that, when non-native listeners are tested, stimuli manipulation using one base token may not guarantee the same results in a study on the cue-use patterns of the stop contrast and the effects of base token need to be confirmed.

It has been known that if a major cue to a phonetic contrast is not clearly realized in a certain environment, listeners can attend to other cues (Colantoni et al., 2015, and references cited therein). In the present study, by orthogonally manipulating naturally produced tokens in several steps of VOT and F0, the cue compositions were perturbed, resulting in many mismatched cue combinations (e.g., [short VOT+low F0], [short VOT+high F0], and [long VOT+mid F0]), which may have confused the listeners' perception of the stop contrast. The set of the other minor cues left in the stimuli may have been perceived stably in this environment. This phenomenon was more apparent in the non- native group, who showed no native-like use of the manipulated VOT and F0 cues and thus possibly more concentration on other constant cues. Despite these mismatched cue combinations, the native group used both VOT and F0 cues systematically and made additional systematic use of other minor cues. The non-native group did not use the manipulated VOT or F0 cue in a native-like way, but showed stable reliance on the minor cues in the context of the mismatched cue combinations. This finding may evidence that listeners maximally use alternative cues to identify a phonetic contrast when major cues sound ambiguous.

The individual listener data for the non-native group showed a wide variety of cue-use patterns (the bottom panels in Figure 4, and Figure 5). Some non-native listeners showed similar patterns to native listeners, while others showed reverse patterns. For example, the lower the F0 value, the more some of the listeners answered as aspirated. These results are consistent with production data in Holliday (2015), a longitudinal study on the learning of Korean stop contrast by native Mandarin Chinese speakers. Some of the participants produced the lenis stops with high F0 values. This error pattern using the F0 cue in signaling the Korean stop contrast does not appear to be resolved even in the perception of the stops for some of the highly advanced learners who participated in this study.

In the present study, the effects of the minor cues left in the stimuli after cue manipulation were considered collectively. It is probable that the minor cues worked collectively in the stop categorization, causing stable response biases toward the selected base token. Given that this study did not seek to investigate the extent to which each minor cue affected perception, a further study is warranted to investigate this issue.

Acknowledgments

An earlier version of this paper was presented at the Spring Conference of the Korean Society of Speech Sciences (June 1, 2019). I am grateful to the audience members, Tae-Yeoub Jang, and three anonymous reviewers for their useful comments; to Sung Yeon Kim for helping recruit participants; and to the subjects for their participation.

References

Abramson, A. S., & Lisker, L. (1985). Relative power of cues: F0 shift versus voice timing. In P. N. Ladefoged & V. A. Fromkin (Eds.), *Phonetic linguistics: Essays in honor of Peter Ladefoged* (pp. 25-33). New York, NY: Academic Press.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Boersma, P., & Weenink, D. (2016). Praat: Doing phonetics by computer (version 6.0.10). [Computer program]. Retrieved from http://www.praat.org/

Colantoni, L., Steele, J., & Escudero, P. (2015). *Second language speech: Theory and practice*. Cambridge, UK: Cambridge University Press.

Holliday, J. J. (2015). A longitudinal study of the second language acquisition of a three-way stop contrast. *Journal of Phonetics*, *50*, 1-14.

Kang, K. H. (2010). Generational differences in the perception of Korean stops. *Phonetics and Speech Sciences*, *2*(3), 3-10.

Kang, Y. (2014). Voice onset time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *Journal of Phonetics*, *45*, 76-90.

Kim, M. R., Beddor, P. S., & Horrocks, J. (2002). The contribution of consonantal and vocalic information to the perception of Korean initial stops. *Journal of Phonetics*, *30*(1), 77-100.

Kong, E. J., & Edwards, J. (2015, August). Individual differences in L2 learners' perceptual cue weighting patterns. *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK.

Lee, H., Politzer-Ahles, S., & Jongman, A. (2013). Speakers of tonal and non-tonal Korean dialects use different cue weightings in the perception of the three-way laryngeal stop contrast. *Journal of Phonetics*, *41*(2), 117-132.

McGuire, G. (2010*). A brief primer on experimental designs for speech perception research* (Master's thesis). University of California at Santa Cruz, California, CA.

Oh, E. (2011). Effects of speaker gender on voice onset time in Korean stops. *Journal of Phonetics*, *39*(1), 59-67.

Oh, E., Idemaru, K., & Kim, B. (2018). The use of a voice onset time cue in the perception of Seoul Korean stops as a function of listener gender. *Korean Journal of Linguistics*, *43*(4), 761-780.

R Development Core Team (2016). *The R project for statistical computing* (version 3.3.2) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Shultz, A. A., Francis, A. L., & Llanos, F. (2012). Differential cue weighting in perception and production of consonant voicing. *The Journal of the Acoustical Society of America*, *132*(2), EL95-EL101.

Silva, D. (2006). Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology*, *23*(2), 287-308.

Winn, M. B., Chatterjee, M., & Idsardi, W. J. (2012). The use of acoustic cues for phonetic identification: Effects of spectral degradation and electric hearing. *Journal of the Acoustical Society of America*, *131*(2), 1465-1479.

• **Eunjin Oh**, Corresponding author
Professor, Dept. of English Language and Literature
Ewha Womans University
52 Ewhayeodae-gil, Seodaemun-gu, Seoul 03760, Korea
Tel: +82-2-3277-3357
Email: ejoh@ewha.ac.kr
Areas of interest: speech production, perception, L2 speech