Cross-generational perception of Korean non-front vowels

Eun Jong Kong[1], Jieun Kang[2]

[1]Korea Aerospace University, [2]Korea University
[1]ekong@kau.ac.kr, [2]nppek@korea.ac.kr

ABSTRACT

This study investigated the perceptual confusion of five non-front vowels /a ʌ o u ɨ/ in Seoul Korean to examine on-going sound changes of /o/-raising. We conducted 5AFC experiments with 25 young adults speaking Seoul Korean in their 20s and 15 older adults above 60s. Auditory stimuli were CV syllables excised from ten young adult female speakers' natural word productions and were presented to the listeners with and without a babbling-noise (+10 SNR). The results revealed generation differences in the accuracy of /u/ identifications where old listeners misheard /u/ for /o/ or /ɨ/ more often than young listeners. In the presence of noise, perceptual confusion existed across the five vowels in a chain-like manner by mishearing the vowels for the next higher or adjacent fronter ones. Regression models predicting accuracies based on stimulus acoustics support that the sound change is not confined to /o/-raising but extended to shifts of /u/ and /ɨ/.

Keywords: sound change, Korean vowel shift, vowel identification, generational difference, vowel raising

1. INTRODUCTION

Production studies for the past decade have reported a sound change of a /o/-raising in Seoul Korean and a subsequent overlap of /o/ with /u/ along the F1 dimension in the acoustic vowel space [4, 5, 16, 19]. That is, the high-back corner of the edges in the vowel space is not occupied by /u/ any more in Seoul Korean but by /o/, which is different from what the acoustic studies in 1990s have documented e.g., [3, 20]. The earliest studies of a sound change further revealed that the tendency was more consistent in females' productions than males', more common in younger speakers' productions than older speakers' and more word-medially than word-initially [4, 5]. Regarding the /o/-raising, there can be several proposals of how widely it would affect the rest of the vowels in the system. Locally, one can view the impact of a /o/-raising exclusively effective on the adjacent vowel /u/ as aforementioned studies referred to the change as a merger [4, 5, 16, 19]. More globally, others can predict that the /o/-raising can affect multiple vowels by pushing and dragging the vowels in the system [2, 9, 10, 11, 14, 15].

The latter proposal of comprehensive shift is supported by recent studies examining speech corpus. [11] examined F1 and F2 of the monophthongs from 30 Seoul speakers aged from teens to 40s using Seoul Corpus [22]. Findings showed that while the F1s of the /o/ vowels were as low as those of the /u/ vowels, the F2 dimension effectively differentiated /o/ from /u/. Moreover, the distance between /ɨ/ and /u/ along the F2 dimension was smaller, as the speakers were younger. Similarly, [10] showed that not just /u/ but /ɨ/ is articulated at more front locations as the speakers are younger based on the spontaneous speech corpus recording Seoul speakers born in from 1930s to 1980s. Together, observations imply a push of /u/ and /ɨ/ toward anterior positions, rendering the prediction of global vowel shifts more convincing than the account of a local merger of /o/-/u/.

Above all, perceptual evidence of distinctive vowels serves as a crucial argument of favoring a global influence over a local effect of /o/-raising in Seoul Korean, because a merger would presuppose a waning contrast between the target vowels. While perception studies of Korean vowels scarcely exist, evidence appears to point to a distinctive perceptual identity of /o/ and /u/. For example, [7] showed that Korean college students were able to identify /o/ and /u/ contrastively with high accuracy rates. [21] also found that F2 played an effective role in young adults' differentiating /o/ from /u/, whereas F1 did not.

Given that none of the existing studies explored global patterns of vowel identifications covering the vowels besides /o/ and /u/, the present research aims to fill this gap by depicting perceptual confusions among the five non-front vowels /a ʌ o u ɨ/ in Seoul Korean. We explore whether perceptual confusions are limited locally to the vowels /o/ and /u/ or they prevail globally across the other non-front vowels /ʌ u ɨ/. For this goal, two variables of listening conditions were considered in the perception experimental design: *Age* of the listeners and *noise* in the signal. The cross-sectional investigation of comparing young (in their 20s) and old (above 60s) listeners' confusions would reveal whether new acoustic properties of /o/ and other non-front vowels present in audio stimuli produced by young speakers are sufficiently different from traditional old forms of vowel acoustics rooted in old listeners' representations (e.g., [6, 8, 13]). The noise variable was deliberately devised to add a difficulty to the perception task. Meaningful differences of identification patterns between sessions with and without noise would reflect unstable representations of the vowels, which are attributable to the on-going sound change.

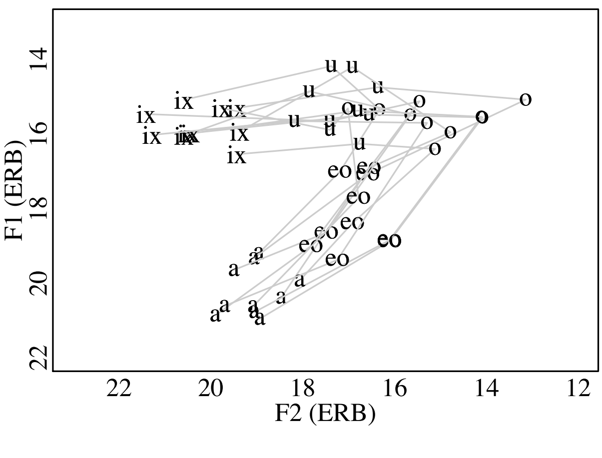
2. PERCEPTION EXPERIMENT

2.1. Methods

2.1.1. Stimuli

The auditory stimulus sets were prepared by recording ten Seoul Korean female speakers in their early 20s. As a subset of larger production study [12], the speakers were selected based on the acoustical properties of /u/ and /o/ vowels, making sure that speakers with a range of /o/-/u/ distance were included in the final set of talkers. From their word productions, the word-final syllables of /CV/ (C= alveolar stops /t t^h t'/; V= five non-front vowels /a ʌ o u ɨ/) were excised to be presented to the listeners, yielding 15 tokens from each speaker (see Fig. 1 for the averaged vowel acoustics in ERB). These multi-talker 150 syllables (15 CVs × 10 talkers) were used in the two experimental blocks: (1) no-noise and (2) noise conditions. The original 150 stimulus tokens were presented in the no-noise condition block, while in the noise condition block, the set of 150 stimuli were combined with babbling noise at the +10dB signal-to-noise ratio for energetic masking.

Figure 1: Distributions of the stimulus tokens averaged across individual speakers (ix=/ɨ/, u=/u/, o=/o/, eo=/ʌ/, a=/a/).



2.1.2. Participants and tasks

25 college students (aged 20-30) and 15 seniors (aged above 60) of Seoul Korean participated in the perception study for a monetary compensation. The young listeners were asked to perform the five-alternative forced-choice (5AFC) task blocked by two noise conditions (NOISE and No-NOISE) programmed in E-prime 2 software [17]. While none of the participants reported any diagnosed hearing problems, some older listeners gave up completing the NOISE block due to a difficulty of detecting signals from a babbling noise. Due to this age-related hearing loss problem, the NOISE block was either not given to the listeners or excluded from the analysis. In the task, the participants were asked to identify the vowel categories upon hearing the CV stimulus over headphones by mouse-clicking one of the five given vowel categories on the monitor. The response vowel categories were displayed in Hangul. The tokens were presented in a randomized order and the two blocks of listening conditions were counter-balanced across the participants of young listeners. It took less than 30 minutes to complete the both sessions.

2.1.3. Analysis

Response category accuracies of 5AFC task were calculated across three experimental conditions (OLD-No-NOISE, YOUNG-No-NOISE, and YOUNG-NOISE) to examine the degrees and directions of perceptual confusion between adjacent vowel categories based on confusion matrices.

In order to investigate the role of acoustic parameters (F1 and F2) in identifying the vowel categories, logistic mixed-effects regression models (using *lme4* package in R platform [1, 18]) were constructed with each of four response category pairs (two adjacent vowels) as a dependent variable: /o/-/u/, /ɨ/-/u/, /o/-/ʌ/, and /ʌ/-/a/. Fixed effect variables were the acoustic parameters of F1 and F2 of the vowels with a random intercept and random slopes of the parameters varying at the listener level. In addition, the experimental conditions were entered as interaction terms with F1 and F2 to assess the effects of age (*AGE model*: OLD-No-NOISE vs. YOUNG-No-NOISE) and listening condition (*NOISE model*: YOUNG-No-NOISE vs. YOUNG-NOISE).

2.2. Results

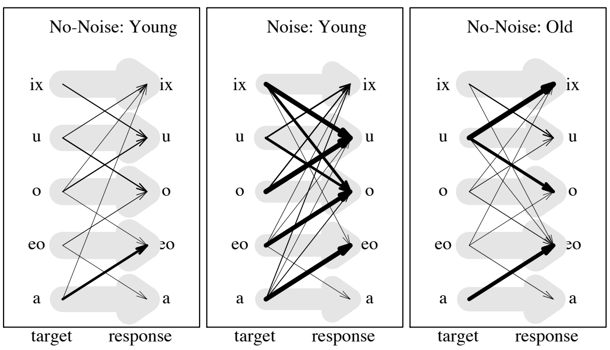
The categorical responses for each of five target vowels were counted to create confusion matrices according to the listening conditions (OLD-No-NOISE, YOUNG-No-NOISE, and YOUNG-NOISE) as in Table 1. While overall accuracy rates were relatively high across the listening conditions, the accuracies were lower in the conditions of YOUNG-NOISE (78.2%), OLD-No-NOISE (88%) and YOUNG-No-NOISE (94.1%), in that order. Examined by vowel types, the vowels [a] and [u] had the two lowest accuracies in the No-NOISE conditions: YOUNG: 90.6%, and 91.8%; OLD: 85.5%, and 70.8%, respectively. In the NOISE condition, the vowel accuracies ranged from 85% (for [u]) to 72.9% (for [a]).

Fig. 2 visualizes the target-response mapping in each listening session, revealing that the responses were mapped to the targets most of the time as illustrated by thick grey arrows and that it was the target [a] that was misheard the most by the listeners across conditions as indicated by the thickest black arrows pointing to [ʌ]. In the YOUNG-NOISE condition, confusion among vowels was sequentially unidirectional. For one, while [a] was often mistaken as [ʌ], it was [o] that [ʌ] was often misheard as. The directions of sequential confusion were consistent in a way that the target non-high vowels were often misheard as the next higher vowels (i.e., [a] as [ʌ], [ʌ] as [o], and [o] as [u]). Unlike the other vowels in the NOISE condition, the vowel [u] was involved in reciprocal confusion with [o] and [ɨ] in that [u] could be confused with [o] or [ɨ] with each other. In the OLD-No-NOISE condition, old listeners were least accurate in identifying the target [u] by perceiving [u] as either [ɨ] or [o].

Table 1: Confusion matrices of five non-front vowels in Seoul Korean separated by the three listening conditions.

	"a"	"ʌ"	"ɨ"	"o"	"u"	total
	(a) YOUNG-No-NOISE					
[a]	680	66	3	0	1	750
[ʌ]	12	719	0	18	1	750
[ɨ]	0	1	720	2	27	750
[o]	1	4	4	712	29	750
[u]	0	1	20	31	698	750
	(b) YOUNG-NOISE					
[a]	574	125	24	22	5	750
[ʌ]	11	595	14	118	12	750
[ɨ]	1	3	547	74	125	750
[o]	2	1	32	580	135	750
[u]	2	1	40	69	638	750
	(c) OLD-No-NOISE					
[a]	385	62	2	1	0	450
[ʌ]	9	430	4	7	0	450
[ɨ]	0	8	423	1	18	450
[o]	1	11	5	423	10	450
[u]	0	3	80	48	319	450

Figure 2: The mappings between target and response vowel categories separated by three conditions. Thickness indicates the quantity of the responses (ix=/ɨ/, u=/u/, o=/o/, eo=/ʌ/, a=/a/).



As summarized in Table 2, the regression models confirmed the age difference in identifying the target [u]. In predicting the responses of [u] and [ɨ] from young and old listeners, the model (AGE [u-ɨ]) yielded significant age-group interactions with F1 and F2 parameters: $\beta_{F1\times AGE}$:-1.911; $\beta_{F2\times AGE}$:-4.121 ($p<.005$). While the tokens with greater F1 and F2 values (implying lower and fronter tongue position) were likely to be heard as [ɨ] over [u], old listeners were less sensitive to both acoustic dimensions than younger listeners as the magnitudes of F1 and F2 beta coefficients were smaller for the old listeners.

Contrary to the model AGE [u-ɨ], the model for the responses [u] and [o] (AGE [u-o]) did not produce a significant age-group interaction with F1 and F2, although higher F1 and lower F2 values (implying lower and more back tongue position) were significant acoustic variables that explained the responses [o] over the responses [u] in both age groups. There were no significant age-group interactions in the models predicting [o] vs. [ʌ] (AGE [o-ʌ]) and [ʌ] vs. [a] (AGE [ʌ-a]). Models of both age groups estimated lower F1 and F2 (implying higher and more back tongue position) to predict [o] from [ʌ] and lower F1 and F2 to predict [ʌ] from [a].

Table 2: Parameter estimations of the logistic mixed effects regression models. Reference levels were YOUNG and No-NOISE in (a) and (b), respectively. Bold indicates p<.05.

(a) AGE	[ɨ-u]	[u-o]	[o-ʌ]	[ʌ-a]
intercept	**-1.79**	**1.125**	**-2.217**	**6.741**
F1	**3.787**	**-1.217**	**-3.962**	**-4.075**
F2	**7.814**	**2.748**	**-2.214**	**-5.265**
Age	**1.910**	**-0.749**	0.267	0.639
F1×Age	**-1.911**	-0.227	0.330	-0.350
F2×Age	**-4.121**	-0.251	0.343	0.440
(b) NOISE	[ɨ-u]	[u-o]	[o-ʌ]	[ʌ-a]
intercept	**-0.857**	**0.492**	**-1.057**	**6.022**
F1	**2.493**	**-1.293**	**-3.710**	**-3.358**
F2	**4.467**	**1.958**	**-1.276**	**-4.954**
Noise	**0.955**	**-1.310**	**2.682**	**-1.620**
F1×Noise	**-1.665**	-0.087	**0.734**	**1.500**
F2×Noise	**-4.571**	**-1.699**	**2.225**	**0.780**

For young listeners, the identifications of the all five vowels significantly differed between NOISE and No-NOISE sessions in terms of sensitivities to F1 and F2. The listeners were likely to identify [ɨ] from [u] when the tokens had greater F1 and F2 values, which indicates a lower and fronter tongue position. This model (NOISE [ɨ-u]) estimated coefficients of F1 and F2 significantly smaller in the NOISE condition than in the No-NOISE condition ($\beta_{F1\times NOISE}$=-1.66; $\beta_{F2\times NOISE}$=-4.56, $p<.001$), indicating that young listeners were less sensitive to the two acoustic dimensions in the presence of distracting noise.

Similarly, the models of responses [o]-[ʌ] (NOISE [o-ʌ]) and [ʌ]-[a] (NOISE [ʌ-a]) had significant noise-condition interactions with F1 and F2 by yielding smaller magnitudes of coefficients of both acoustic parameters in the NOISE sessions. For example, in the model NOISE [o-ʌ], the magnitude of F2 coefficient was smaller in the NOISE session that the parameter was estimated to be not significant (NOISE: β_{F2}=-0.16, p=.3; NoNOISE: β_{F2}=-1.27, p<.001). For the responses [ʌ] and [a] (NOISE [ʌ-a]), lower F1 and F2 (implying higher and more back tongue position) predicted [ʌ] over [a] in both listening sessions but the listeners' reliance on F1 and F2 was significantly mitigated in the NOISE condition: $\beta_{F1\times NOISE}$=1.50; $\beta_{F2\times NOISE}$=.78 (p<.05).

Finally, it is noted that there was no significant coefficient difference between the noise-conditions in the F1 parameter of the model of [u]-[o] responses (NOISE [u-o]: $\beta_{F1\times NOISE}$=.087, p=.67), while F1 was a significant parameter in predicting [u] from [o] in both noise conditions (NOISE: β_{F1}=-1.24, p<.0001; NoNOISE: β_{F1}=-1.34, p<.001), indicating that lower F1 (higher tongue position) predicted [u] responses over [o]. The tendency that a presence of noise did not distract the young listeners' perceptual reliance on F1 but did on F2 seems to suggest that F2 is a more important perceptual dimension than F1 in identifying [o] from [u], primarily taking up listeners' attentional resources.

3. DISCUSSION

The present study investigated young and old adult listeners' identification patterns of five vowels /a ʌ o u ɨ/ in Seoul Korean, aiming to examine the status and nature of the on-going vowel sound change. While a /o/-raising could potentially lead to a merge of /o/ and /u/, that does not seem to be the final destinations of the vowel change in Seoul Korean. Identification error patterns revealed that the two vowels /o/ and /u/ maintained highly distinct with each other in perception across age-groups and importantly they were not mutually confusable: Old listeners tended to mistake /u/ for /o/, but the opposite errors were few.

Rather, the current findings support that the /o/-raising subsequently pushed /u/ to a fronter position and further pushed and dragged /ɨ/ down to a lower position in the vowel space. The generational difference in the acoustic representations of /u/, /o/ and /ɨ/ appears to indicate the shift of /u/ to /ɨ/ in progress. The old listeners' conservative ears tended to identify young adults' realizations of /u/ as /ɨ/ with less sensitivities to acoustic dimensions of both F1 and F2 compared to the young listeners. Contrary to the responses /u/ and /ɨ/, the generational discrepancy in utilizing the acoustic dimensions was not observed in the vowels /o/ and /u/, although older listeners produced relatively more errors of perceiving [u] as [o] than younger listeners. Given that the Korean /o/-raising has been reported to be in progress for a decade, the current findings can possibly suggest that the innovative acoustic characteristics of /o/ in Seoul Korean has been sufficiently transmitted throughout the speech communities across generations. Taken together with existing production studies reporting lowered /ɨ/ and fronted /u/ in Seoul Korean, the present perceptual evidence suggests that vowels /o/, /u/, and /ɨ/ are sequentially undergoing changes in terms of their acoustic characterization.

Although not confirmed by the cross-generational model of acoustic parameters, the most frequent errors of /ʌ/ for /a/ suggest that the two vowels /a/ and /ʌ/ might be involved in the on-going sound change of Seoul Korean. The misperception of /a/ as /ʌ/ stood out among the back vowels by its highest error rate even in the no-noise condition by the young listeners. This pattern of perceptual confusion is consistent with the observation in production study [11] where the distance between /a/ and /ʌ/ was closer in the young (teens and 20s) speakers' productions than in the older (40s) speakers'. The misidentification of /a/ for /ʌ/ might come from a mismatch between existing perceptual representation and innovative forms of production, i.e., current vowel stimuli. Given that the confusions were found both in young and old listeners, the change related to /a/ and /ʌ/ might be recent one that follows after the /o/-raising and /u/ and /ɨ/-fronting. Taken together, whether weak or strong, the confusion patterns provided perceptual evidence of a sound change in the non-front vowels in Seoul Korean.

4. CONCLUSION

To conclude, the present paper provides perceptual evidence of the vowel sound change in Seoul Korean, which can be characterized as successive shifts of /o/-raising, /u/-fronting, and /ɨ/-fronting & lowering. Although weak, evidence of change in the vowels /ʌ/ and /a/ is taking place in Seoul Korean. Discussions on motivations and phonological consequences of the sound changes should follow to comprehensively understand the vowel system of Seoul Korean.

5. ACKNOWLEDGEMENTS

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2017S1A5A2A01024268).

6. REFERENCES

[1] Bates, D., Maechler, M., Bolker, B., Walker, S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48.

[2] Cho, S. 2003. An acoustic study of Korean vowel system. *Journal of the International Network for Korean Language and Culture*, 24, 427-441.

[3] Chung, I. 1997. Generational differences of monophthongs in the standard Korean: Experimental approach, *Malsori*, 33, 111-125.

[4] Han, J., Kang, H. 2013. Cross-generational Change of /o/ and /u/ in Seoul Korean I: Proximity in Vowel Space. *Phonetics and Speech Sciences*, 5(2), 25-31.

[5] Han, J., Kang, H., Kim, J. 2013. Cross-Generational Differences of /o/ and /u/ in informal text reading. *Phonetics and Speech Sciences*, 5(4), 201-207.

[6] Harrington, J., Kleber, F., Reubold, U. 2008. Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study. *J. Acoust. Soc. Am*., 123(5), 2825-2835.

[7] Igeta, T., Sonu, M., Arai, T. 2014. Sound change of /o/ in modern Seoul Korean: Focused on relations with acoustic characteristics and perception. *Phonetics and Speech Sciences*, 6(3), 109-119.

[8] Jacewicz, E., Fox, R. A. 2012. The effects of cross-generational and cross-dialectal variation on vowel identification and classification. *J. Acoust. Soc. Am*., *131*(2), 1413-1433.

[9] Jang, H., Shin, J., Nam, H. 2015. Aspects of vowels by ages in Seoul dialect. *Studies in Phonetics, Phonology and Morphology*, 21(2), 341-358.

[10] Kang, Y. 2014. A corpus-based study of positional variation in Seoul Korean vowels. *Japanese/Korean Linguistics*, 23, 1-20.

[11] Kang, E. J., Kong, E. 2016. Static and dynamic spectral properties of the monophthong vowels in Seoul Korean: Implication on sound change. *Phonetics and Speech Sciences*, 8(4), 39-47.

[12] Kong, E. J., Kang, S., Seo, M. 2014. Gender difference in the affricate productions of young Seoul Korean speakers. *J. Acoust. Soc. Am*., 136(4), EL329-EL335.

[13] Labov, W. 2001. *Principles of Linguistic Change. II: Social Factors.* Blackwell, Oxford

[14] Lee, J., Yoon, K., Byun, K. 2016. A study of vowel shift in Seoul Korean. *The Journal of Studies in Language*, 31(4), 979-998.

[15] Lee, H., Shin, W., Shin, J. 2017. A sociophonetic study on high/mid back vowels in Korean. *Phonetics and Speech Sciences*, 9(2), 39-51.

[16] Moon, S. 2007. A fundamental phonetic investigation of Korean monophthongs. *Malsori*, 62, 1-18.

[17] Psychology Software Tools, Incorporated. [E-Prime 2.0]. 2012. http://www.pstnet.com

[18] R Core Team 2017. *R: A language and environment for statistical computing. R Foundation for Statistical Computing*, Vienna, Austria. https://www.R-project.org/.

[19] Seong, C. 2004. An acoustic analysis on the Korean 8 Monophthongs: with respect to the acoustic variables on the F1/F2 vowel space. *J. Acoust. Soc. Am*., 23, 454-461.

[20] Yang, B. 1996. A comparative study of American English and Korean vowels produced by male and female speakers. *Journal of Phonetics*, 24, 245-261.

[21] Yun, J., Seong, C. 2013. Effect of F1/F2 manipulation on the perception of Korean vowels /o/ and /u/. *Phonetics and Speech Sciences*, 5(3), 39-46.

[22] Yun, W., Yoon, K., Park, S., Lee, J., Cho, S., Kang, D., Byun, K., Hahn, H., Kim, J. 2015. The Korean corpus of spontaneous speech. *Phonetics and Speech Sciences*, 7(2), 103-109.