Spring 2021 June 6, 2021
LINGUIST 245B Sarang Jeong

The relation between perception and production in Korean three-way stop contrast

- A pilot experiment on the perception of the younger age group -

1. Introduction

The purpose of this study is to investigate whether individual language users' perception is parallel to their production when their language community is going through a sound change. When a sound change is in progress, do speakers with the new and old variants have perceptual maps that reflect their own respective variants? Or do they share similar perceptual maps despite having different variants in production? I will address this question by examining Korean stop contrast whose primary cue has been shifting from VOT to F0. I hypothesize that there will be discrepancy between listeners' perceptual maps and their production patterns. Through an auditory identification experiment, this preliminary study explores one of the predictions made by the hypothesis, which relates to the perception of the younger age group.

1.1. Background

Relation between production and perception

The relation between production and perception has been a central issue in phonetic theories and theories of speech perception. For successful communication, there needs to be a sufficient similarity between the speech signal produced by the speaker and the linguistic unit perceived by the listener. Some phonetic theories even make claims about the production-perception relationship *within* one individual language user (Beddor 2015). For instance, Motor Theory claims that the listener refers to his/her own motor system in order to perceive speech (Galantucci et al. 2006, Liberman & Mattingly 1985). Similarly, exemplar theory often assumes that an individual's production forms are derived from exemplar clouds consisting of all the perceptual inputs that the language user has ever received (Kirchner 2012, Pierrehumbert 2001).

The link between individual language user's production and perception is also of great importance in the study of sound change. For a sound change to take place, listeners need to *produce* the form they have *perceived* as salient (Beddor 2012, Harrington et al. 2008, Lindblom et al. 1995, Ohala 1981), either by misparsing signal (Ohala 1993, Ogura 2012) or by paying attention to reduced forms (Beddor 2009, Bybee 2012, Harrington et al. 2015, Lindblom et al. 1995). Indeed, there is a plenty of empirical evidence for symmetry between production and perception of individual language users under ongoing sound changes (Coetzee et al 2018, Beddor et al. 2018, Kirby 2014, Beddor et al. 2015).

However, perception needs to be more flexible than production (e.g. Schertz et al. 2019, Beddor 2015) because the listener has to understand the speech signal in the face of different types of noise including speaker variation. Indeed, a large body of research has demonstrated that listeners are sensitive to fine-grained phonetic variation (e.g. McMurray et al. 2002, Clayards et al. 2008).

The importance of perceptual flexibility becomes especially prominent in the context of sound change in progress, for a language community undergoing a sound change inevitably includes groups of speakers that systematically differ in their production. From this context my research question

arises: Under an ongoing sound change, do individual language users have perceptual maps that closely correspond to their own production patterns, or do language users differ in production but have more similarity in perception?

I will investigate this question by exploring Korean three-way stop contrast. The Korean data provides a good place to explore the question in hand for two reasons. First, it is currently undergoing a sound change, which allows for direct experiments with speakers rather than relying on previous studies to indirectly track the course of sound change. Second, the given sound change has been reported to be a production-driven change, led by lenition bias in VOT of high-frequency words (Bang et al. 2018). Thus, it will be an ideal case to examine whether perception changes in tandem with production. In the next subsection, I will give a linguistic description of Korean stop contrast and the sound change it has been going through, and briefly review previous works on the production and perception of the stop contrast.

Korean stop contrast

Korean stops are "typologically unusual" because they are all voiceless word-initially yet have a three-way distinction (Cho et al. 2002, 193). The three categories are labeled as lenis/lax (C), fortis/tense (C*), and aspirated (C^h). The triplets have three places of articulation: bilabial, alveolar, and velar (see Table 1 – 2). The contrast has various phonetic cues, but the VOT of the stop and the F0 of the following vowel are the primary cues. VOT is longer for the aspirated and the lenis stop and shorter for the fortis stop (Bang et al. 2018). F0 of the following vowel is lower for the lenis stop and higher for the aspirated stop, the fortis stop taking up the intermediate place in F0 continuum (Lee & Jongman 2012, Bang et al. 2018, Kong et al. 2011).

Table 1. Stops in Korean: lenis, fortis, and aspirated

	Bilabial	Alveolar	Velar
Lenis	p	t	k
Tense	p*	t*	k*
Aspirated	p^h	t^h	k^h

Table 2. Minimal triplets for Korean stops in word-initial position

Lenis		Tense		Aspirated	
paŋ	'room'	p*aŋ	'bread'	p^haŋ	'bang'
tal	'moon'	t*al	'daughter'	t^hal	'mask'
kol	'valley'	k*ol	'shape'	k^hol	'call'

The given contrast is going through a sound change. Traditionally, the primary cue to Korean stop contrast was attributed to VOT: the aspirated stop had the longest VOT and the tense stop the shortest VOT, the lenis stop being the intermediate one. The F0 of the following vowel was merely a redundant phonetic property that was physiologically caused by the laryngeal contrast (Han & Weitzman 1967; Hardcastle 1973; Kang & Han 2013; Kim 1965). However, recent studies reported a merger in VOT and an enhancement of F0 contrast for the lenis (lower F0) *vs*. aspirated (higher F0) distinction in phrase-initial position (Beckman et al. 2014, Silva 2002, Kim et al. 2002, Kong et al. 2011, Lee et al. 2013). Through an apparent-time study of a spoken corpus, Bang et al. (2018) discovered that both VOT contrast reduction and F0 contrast enhancement are progressing in parallel: The younger the speaker was, he/she relied more on F0 and less on VOT to distinguish lenis and aspirated stops.

On the production part, lenis *vs*. aspirated contrast has fully become an F0 contrast for younger

speakers. For example, Kim et al. (2019) showed that adult speakers contrasted the three stops along the dimension of F0 in every focus environment (Figure 1, left side). The VOT for the aspirated and lenis stops were not different, and only the tense stop had significantly shorter VOT (Figure 1, right side). A number of other production studies reported similar findings (e.g. Bang et al. 2018, Cho et al. 2002, Lee & Jongman 2012).

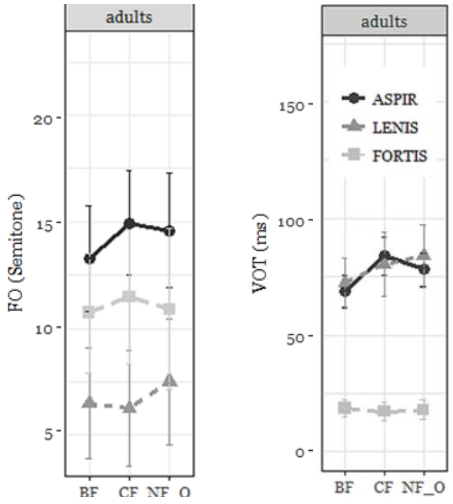

Figure 1. F0 (left) and VOT (right) values for each consonant.
BF = broad focus, CF = contrastive focus, NF_O = narrow focus as object.
(Modified from Kim et al. 2019, 1517-1518)

Several studies investigated the perception aspect of the given contrast. One of the most comprehensive studies was done by Lee et al. (2013). They tested Seoul Korean and Kyungsang Korean speakers' perception of bilabial stops. Overall, an extremely short VOT value indicated the tense stop, while low F0 and high F0 indicated the lenis stop and aspirated stop, respectively. In addition, an extremely long VOT indicated the aspirated stop, regardless of the pitch. Other perceptual studies presented comparable findings (e.g. Kim et al. 2002, Kim 2004, Kong et al. 2011).

Despite the large number of studies focusing on either production or perception side of the given sound change, little work is done that examines both sides at the same time. Two such works are Kong et al. (2011) and Schertz et al. (2019), which compare perception and production of the given contrast and concludes perception is more variable than production. Although these studies suggest that language users' perception is different from their production of the contrast, they either lack direct comparison of perception and production by Seoul Korean (the dialect where the sound change has occurred) speakers (Schertz et al. 2019), or do not include language users that presumably have the old variants (Kong et al. 2011). Also, Kong et al. (2011) uses the recordings from the production experiment as the stimuli for the perception experiment, which has the disadvantage of having gaps in the stimuli for combinations of F0 and VOT values, thus disallowing systematic analysis of the cues. Directly comparing the production and perception of individual language users, including both younger and older speaker groups, will provide a more complete insight into the relationship between production and perception under a sound change.

1.2. Hypothesis and predictions

The hypothesis of the current study is that there will be discrepancy between listeners' perceptual maps and their production patterns of the stop contrast in favor of similarity between perceptual maps

of younger and older speakers. The hypothesis makes following predictions:

1) In production, the younger speakers will show similar VOT values for the lenis and aspirated stops, distinguishing them with high F0 value for the aspirated and low F0 value for the lenis. They will have intermediate F0 value and shorter VOT value for the tense stop (Figure 2a).
2) In production, the older speakers will show varying VOT values for the stops, with incremental values for the tense, lenis, and aspirated stops. They will have similar F0 values for the tense and aspirated stops, and lower F0 value for the lenis stop (Figure 2b).
3) In perception, the younger speakers will have category boundary between the lenis and aspirated stops along the F0 dimension, but the boundary will be tilted so that longer VOT conditions more aspirated responses and shorter VOT conditions more lenis responses (Figure 3a).
4) In perception, the older speakers will have category boundary between the lenis and aspirated stops along the VOT dimension, but the boundary will be tilted so that higher F0 conditions more aspirated responses and lower F0 conditions more lenis responses (Figure 3b).[1]

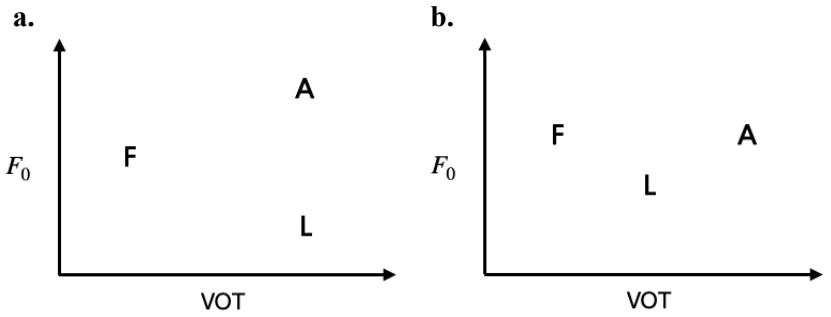

Figure 2. Predicted production patterns of the younger group (a) and older (b) groups

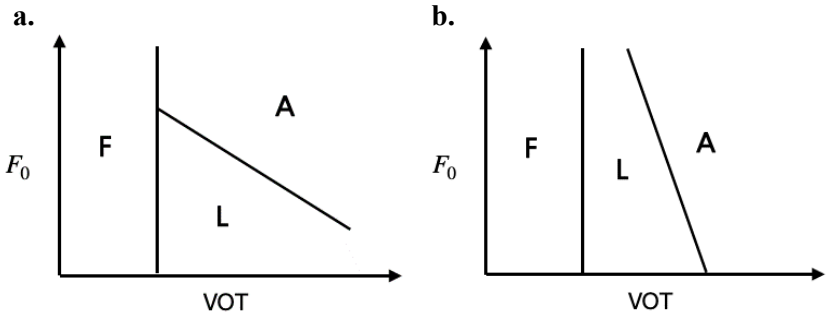

Figure 3. Perceptual maps of the younger (a) and older (b) groups predicted by the hypothesis

Predictions 1 and 2, which are about production patterns, are expected to hold whether the hypothesis is correct or not. In contrast, the hypothesis will be rejected if predictions 5 and 6 turn out to be true.

5) In perception, the younger speakers will have category boundary between the lenis and aspirated stops along the F0 dimension only (Figure 4a).
6) In perception, the older speakers will have category boundary between the lenis and aspirated stops along the VOT dimension only (Figure 4b).

[1] Note that the perceptual map for the fortis stop is predicted to be short VOT range constantly, following previous research's findings (e.g. Kong et al. 2011).

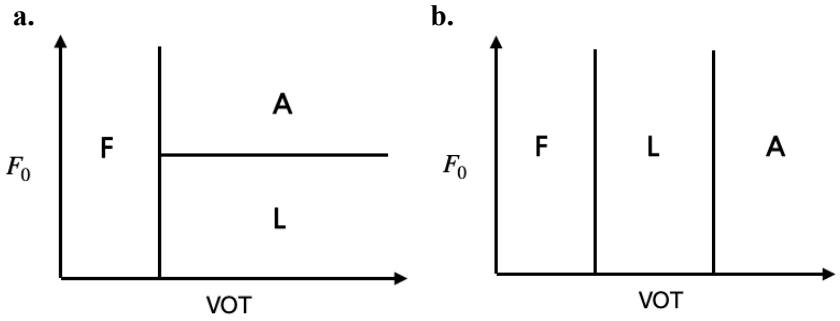

Figure 4. Perceptual maps of the younger (a) and older (b) groups that will reject the hypothesis

In the present paper, only the prediction regarding younger age group's perception will be investigated through experiment. The production patterns depicted in Figure 2 will be assumed to be true. Experimental results will be compared with the Prediction 3 (Figure 3a) and Prediction 5 (Figure 4a), each of which will support or reject the hypothesis. Experimental materials, data and analysis script can be found at: https://github.com/sarangjeong/korean-stops. The experiment was preregistered on OSF (https://osf.io/wzn6b).

2. Experiment

2.1. Methods

Participants

Twenty Korean native speakers in their 20s were recruited through Prolific. The participants who indicated that either they or their parents speak Kyungsang dialect were excluded due to a possible confound resulting from tonal contrast in the dialect. The final number of participants was 17 (female = 9, male = 8, mean age = 25.12).

Materials

As base tokens, a Seoul Korean male speaker (25 years old) produced the aspirated (*pʰaŋ*) and tense (*p*aŋ*) variants of a minimal triplet. The combination of sex and age of the speaker provides the most appropriate voice for testing varying VOT and F0 values because younger speakers tend to have the new F0 distinction, but males are behind females in the transition to complete F0 distinction.

The base tokens were manipulated to create stimuli varying in VOT and F0 values, which allowed for controlling unintended variables (e.g. duration, pitch accent) interfering with results. The VOT range for manipulation was 10–115 ms, with 8 steps of 15 ms. The F0 range for manipulation was 120–190 Hz, with 8 steps of 10 Hz. The ranges were based on the actual range of the given speaker's production. Manipulation was done using a Praat script (Winn 2020). The mean intensity of each stimulus was increased to 70 dB using Praat (Boersma & Weenink 2021). A total of 64 stimuli were created (8 steps of F0 * 8 steps of VOT).

Procedure

The experiment was a forced-choice identification task. Each participant listened to a stimulus and identified it by clicking one of the three options on the screen, which was a labial minimal triplet (*paŋ, p*aŋ, pʰaŋ*) in Korean orthography in a fixed order. The same set of stimuli was repeated three times, presented in a random order in each time (total: 64*3=192). Participants could take a break in between sets. Before the actual trial phase, there were an audio check phase and a familiarization phase. In the audio check phase, the participant listened to three Korean sentences and answered a question about

each sentence. When the participant chose a wrong answer two times or more, the experiment did not proceed. In the familiarization phase, the participant identified 10 stimuli in the same way as in the actual trials. The familiarization stimuli were always given in the same order after being randomly selected from the actual stimuli. The experiment took 11.7 minutes on average. Each participant was paid 3.30 GBP.

2.2. Results

Mean response rates for each category (aspirated, lenis, and tense) across F0 and VOT continua are each shown in Figure 5–7. Mean response rates of all categories across F0 and VOT continua are given in Figure 8. The steps of F0 and VOT in the plots are normalized as 1–8. The question of interest is whether young speakers have category boundary between aspirated and lenis not only as a function of F0 but also as a function of VOT.

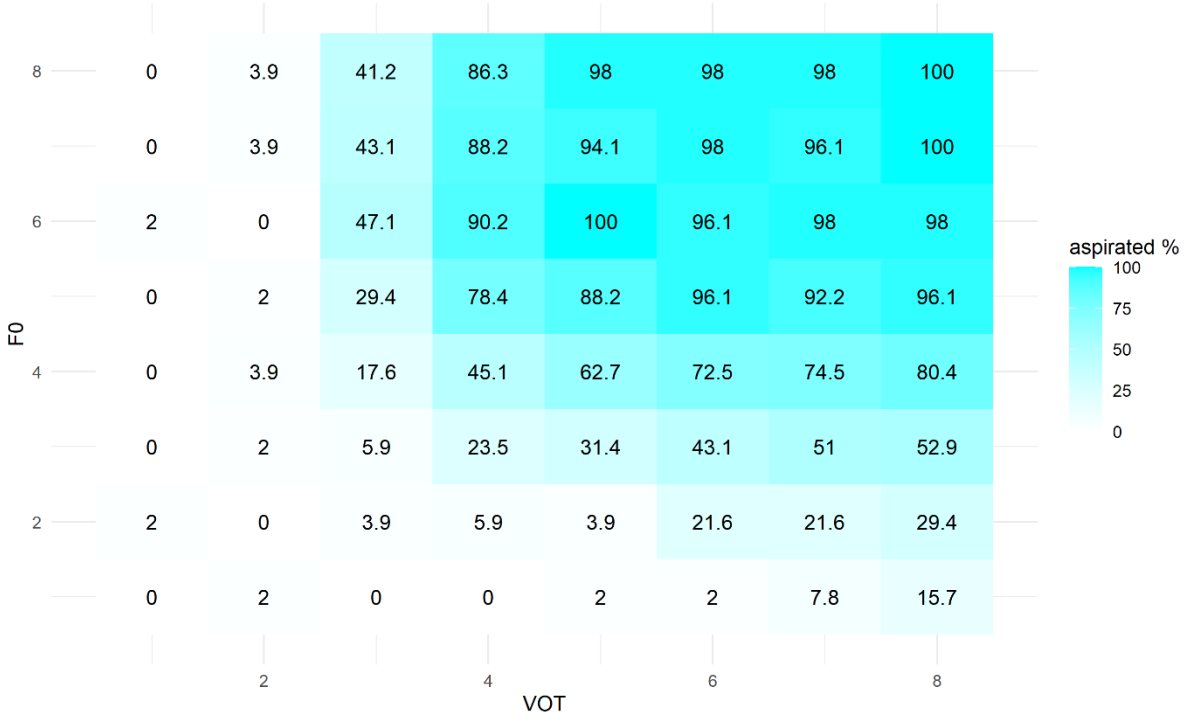

Figure 5. Mean aspirated response rate (%) of the younger age group across VOT and F0

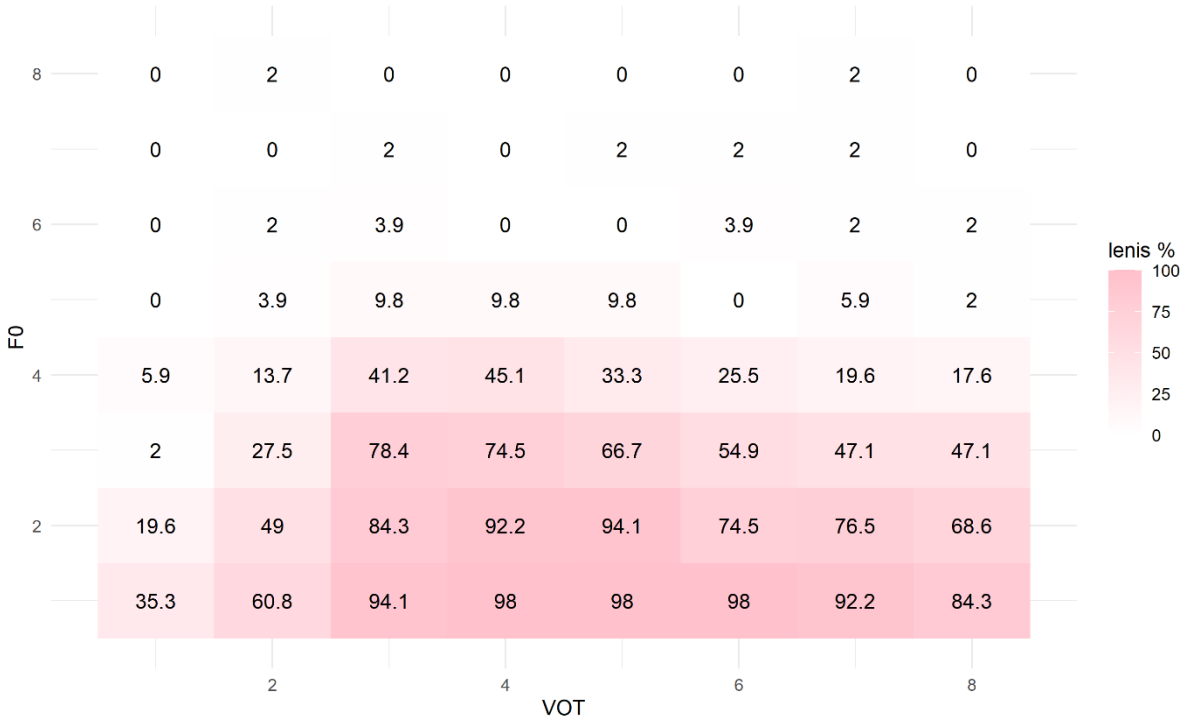

Figure 6. Mean lenis response rate (%) of the younger age group across VOT and F0

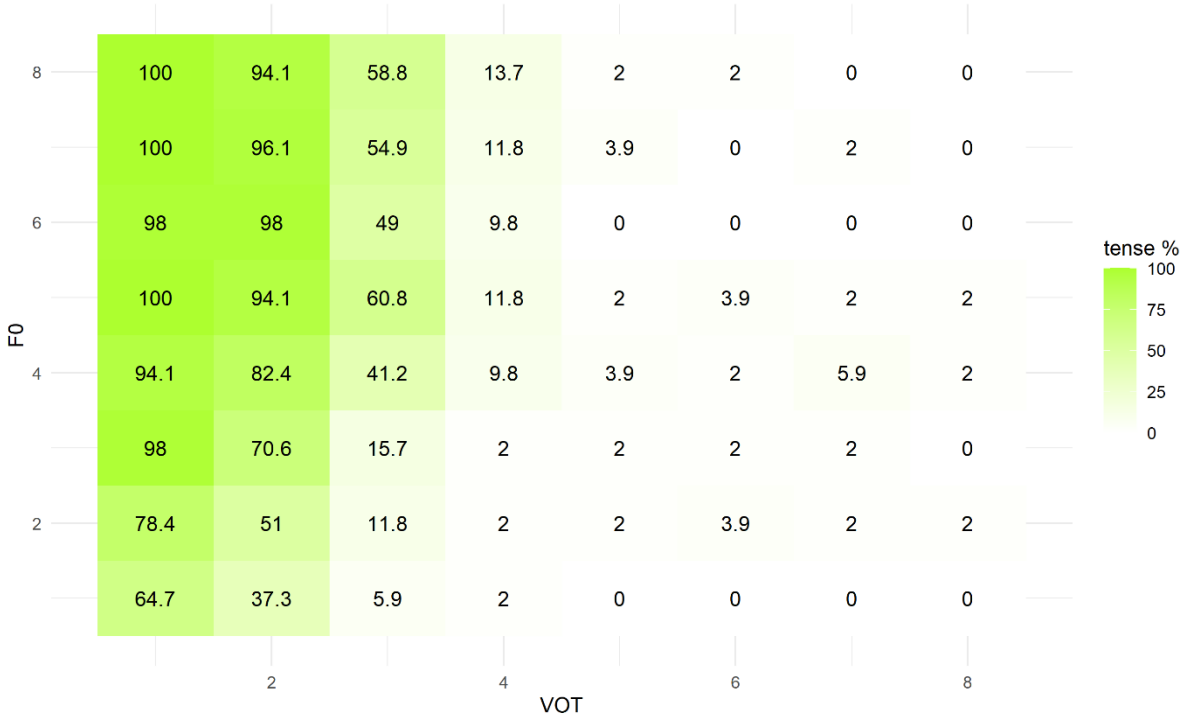

Figure 7. Mean tense response rate (%) of the younger age group across VOT and F0

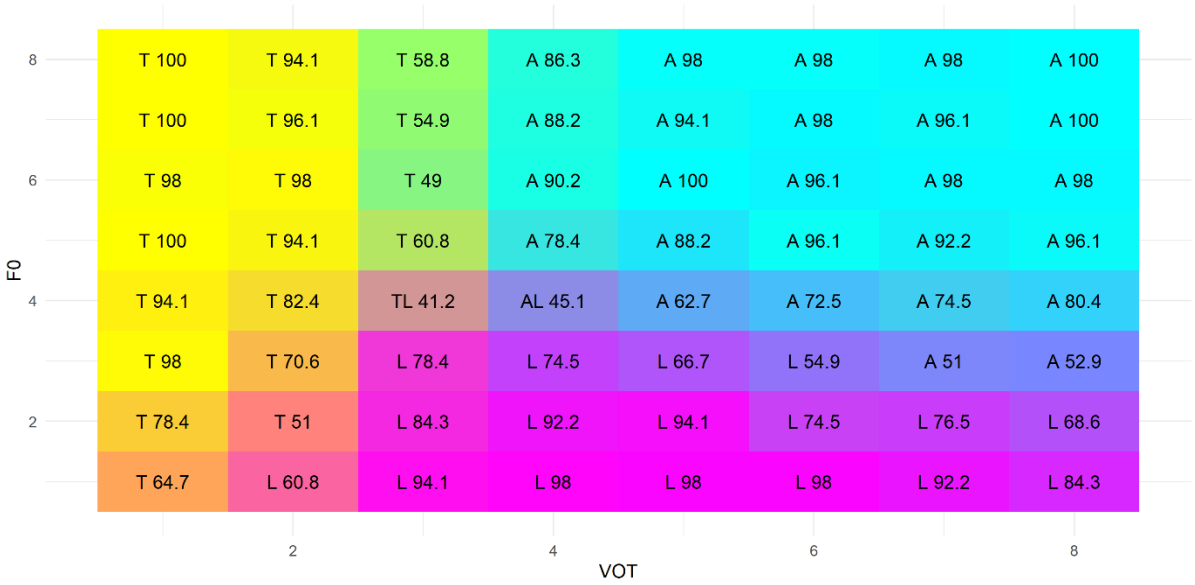

Figure 8. Mean response rates for each category (%) of the younger age group across VOT and F0 (A = aspirated, L = lenis, T = tense)

The following observations were made from the plots. First, aspirated responses are associated with higher F0 and longer VOT, lenis with lower F0 and longer VOT, and tense with shorter VOT. Second, the boundary between aspirated and lenis is associated with not only F0 but also VOT such that F0 being fixed, longer VOT is associated with more aspirated responses. This is in congruence with the prediction of the hypothesis that the boundary between aspirated and lenis will reflect both F0 and VOT rather than being a strict function of F0 alone. Third, tense category is associated with F0 such that VOT being fixed, higher F0 is associated with more tense responses.

These observations were statistically confirmed. I conducted three separate mixed-effects logistic regression analyses for each response category. Each model predicted aspirated, lenis, or tense response over the other responses from fixed effects of F0 (1 through 8, centered and rescaled) and VOT (1 through 8, centered and rescaled), and the interaction between the two.[2] Each model included the maximal random effects structure that allowed the model to converge: by-participant random intercepts as well as by-participant random slopes for the fixed effects of F0 and VOT. For the lenis model, by-participant random slopes for the interaction between F0 and VOT were included as well.

The model predicting aspirated responses revealed a main effect of F0 such that listeners were more likely to identify a stimulus as aspirated when the F0 value of the stimulus was higher (β = 2.53, *SE* = .19, $p < .0001$), and a main effect of VOT such that listeners were more likely to identify a stimulus as aspirated when the VOT value was higher (β = 2.84, *SE* = .23, $p < .0001$). The model also revealed a positive interaction between F0 and VOT such that listeners were even more likely to give an aspirated response when F0 and VOT increased simultaneously (β = 1.94, *SE* = .12, $p = < .0001$).

The model predicting lenis responses revealed a negative main effect of F0 such that listeners were less likely to identify a stimulus as lenis when the F0 value of the stimulus was higher (β = -3.49, *SE* = .25, $p < .0001$). I also observed a negative interaction between F0 and VOT such that listeners were less likely to identify a stimulus as lenis as F0 and VOT increased in tandem (β = -1.17, *SE* = .20, $p < .0001$). A main effect of VOT alone was insignificant.

[2] All mixed-effects analyses reported in this paper were conducted with the `lme4` package (Bates et al. 2015) in R (R Core Team, 2020).

Finally, the model predicting tense responses revealed a main effect of F0 such that listeners were more likely to identify a stimulus when the F0 of the stimulus was higher (β = .68, *SE* = .21, $p < .001$), and a negative main effect of VOT such that listeners were less likely to give a tense response when the VOT value of the stimulus was higher (β = -6.15, *SE* = .62, $p < .0001$). The size of the VOT effect was larger than that of the F0 effect, in accordance with the observation that VOT is the main determiner of tense category. I also observed a negative interaction between F0 and VOT such that listeners were less likely to give a tense response when the F0 and VOT of the stimulus increased together (β = -1.00, *SE* = .17, $p < .0001$).

3. Discussion

This study was motivated by the following question: Does the perception of a speaker closely mirror their own production patterns or does it reflect the production patterns of the language community as a whole when there is an ongoing sound change? Specifically, the present experiment examines the perceptual map of the younger speakers with regard to Korean three-way stop contrast, whose primary cue has been shifting from VOT to F0. The hypothesis was that younger speakers will use both F0 and VOT to distinguish aspirated and lenis stops in perception even though they only use F0 to contrast aspirated and lenis stops in production. I conducted an online identification task with Seoul Korean speakers in their 20s, where they identified stimuli varying in F0 and VOT. A mixed-effects logistic regression analysis was conducted for each response category. For aspirated identification, F0 and VOT, as well as their interaction, had significant positive effects. For lenis identification, F0 and an interaction between F0 and VOT had significant negative effects. Finally, for tense identification, a negative main effect of VOT and a positive main effect of F0 were found, along with a negative interaction between the two.

The prediction of the hypothesis was borne out. Namely, the category boundary between aspirated and lenis stops was indeed a function of F0 *and* VOT for younger listeners. The significant predictors of the aspirated outcome were F0 and VOT, as well as their interaction, all in the positive direction. For lenis responses, the significant predictors were F0 (negative) and the interaction between F0 and VOT (negative). It is notable that a main effect of VOT was not significant, for short VOT was the primary cue for the lenis stop before the sound change. The negative effect of longer VOT was observable only in the interaction with higher F0.

What was not predicted by the hypothesis but was observed in the experimental results is that F0 is also a significant predictor of tense identification. Although tense stops occupy the intermediate position in the F0 dimension of young Korean speakers' production (Kim et al. 2019), I did not expect a significant effect of F0 on tense responses because short VOT is the primary cue for the tense category both before and after the sound change. However, a regression analysis revealed that, although smaller in magnitude than VOT, F0 did have a significant positive effect on tense responses.

The results suggest that under a sound change, the perceptual map of speakers change in such a way that a category *expands* to encompass both the old and new variants, instead of altogether shifting the dimension of the category from one to another. The hypothesis of the present study will gain more support by examining production as well as perception of individual speakers. Extending stimuli to the other places of articulation will also allow us to generalize the results to all Korean stops. Most importantly, conducting an experiment with the older age group is crucial for a more comprehensive understanding of the relation between perception and production under a sound change. It will be particularly intriguing to see if the older age group, who retain the old variant in their production, have a similar perceptual map as the younger age group. These will be the tasks of follow-up experiments.

References

Bang, H. Y., Sonderegger, M., Kang, Y., Clayards, M., & Yoon, T. J. (2018). The emergence, progress, and impact of sound change in progress in Seoul Korean: Implications for mechanisms of tonogenesis. *Journal of Phonetics*, *66*, 120-144.

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

Beckman, M. E., Li, F., Kong, E. J., & Edwards, J. (2014). Aligning the timelines of phonological acquisition and change. *Laboratory phonology*, *5*(1), 151-194.

Beddor, P. S. (2009). A coarticulatory path to sound change. *Language*, *85*(4), 785–821.

Beddor, P. S. (2012). Perception grammars and sound change. *The initiation of sound change: Perception, production, and social factors*, 37-55.

Beddor, P. S. (2015). The relation between language users' perception and production repertoires. In *ICPhS*.

Beddor, P. S., Coetzee, A. W., Styler, W., McGowan, K. B., & Boland, J. E. (2018). The time course of individuals' perception of coarticulatory information is linked to their production: Implications for sound change. *Language*, *94*(4), 931-968.

Boersma, P., & Weenink, D. (2021). Praat: doing phonetics by computer [Computer program]. Version 6.1.48. http://www.praat.org/.

Bybee, J. L. (2012). Patterns of lexical diffusion and articulatory motivation for sound change. In M.-J. Solé & D. Recasens (Eds.), *The initiation of sound change, perception, production, and social factors* (pp. 210–234). Amsterdam, the Netherlands: Benjamins.

Cho, T., Jun, S. A., & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of phonetics*, *30* (2), 193-228.

Clayards, M., Tanenhaus, M. K., Aslin, R. N., & Jacobs, R. A. (2008). Perception of speech reflects optimal use of probabilistic speech cues. *Cognition*, *108* (3), 804-809.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic bulletin & review*, *13*(3), 361-377.

Han, M. S., & Weitzman, R. S. (1967). *Acoustic features in the manner-differentiation of Korean stop consonants* (Vol. 5). Acoustic Phonetics Research Laboratory, University of Southern California.

Hardcastle, W. J. (1973). Some observations on the tense-lax distinction in initial stops in Korean. *Journal of phonetics*, *1*(3), 263-272.

Harrington, J., Kleber, F., & Reubold, U. (2008). Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study. *The Journal of the Acoustical Society of America*, *123*(5), 2825-2835.

Harrington, J., Kleber, F., Reubold, U., & Siddins, J. (2015). The relationship between prosodic

weakening and sound change: Evidence from the German tense/lax contrast. *Laboratory Phonology*, *6*(1), 87–117.

Kang, Y., & Han, S. (2013). Tonogenesis in early Contemporary Seoul Korean: A longitudinal case study. *Lingua*, *134*, 62-74.

Kim, C. W. (1965). On the autonomy of the tensity feature in stop classification (with special reference to Korean stops). *Word*, *21*(3), 339-359.

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.

Kim, S., Yang, A., Chen, A., & Cho, T. (2019). Developmental pathway of phonetic fine-tuning of phonological contrast in reference to information structure: The case of three-way contrastive stops in Korean. In *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 1515-1518). Australasian Speech Science and Technology Association Inc.

Kirby, J. P. (2014). Incipient tonogenesis in Phnom Penh Khmer: Acoustic and perceptual studies. *Journal of Phonetics*, *43*, 69-85.

Kirchner, R., & Moore, R. K. (2012). Modeling exemplar-based phonologization. *The Oxford handbook of laboratory phonology*, 332-344.

Kong, E. J., Beckman, M. E., & Edwards, J. (2011). Why are Korean tense stops acquired so early?: The role of acoustic properties. *Journal of Phonetics*, *39*(2), 196-211.

Lee, H., & Jongman, A. (2012). Effects of tone on the three-way laryngeal distinction in Korean: An acoustic and aerodynamic comparison of the Seoul and South Kyungsang dialects. *Journal of the International Phonetic Association*, *42*(2), 145-169.

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.

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1-36.

Lindblom, B., Guion, S. G., Hura, S., Moon, S.-J., & Willerman, R. (1995). Is sound change adaptive? *Rivista di Linguistica*, *7*, 5–36.

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within category phonetic variation on lexical access. *Cognition*, *86* (2), B33-B42.

Ogura, M. (2012). The Timing of Language Change. *The handbook of historical sociolinguistics*, *68*, 427.

Ohala, J. J. (1993). Sound change as nature's speech perception experiment. *Speech Communication*, *13*(1-2), 155-161.

Ohala, J., Masek, C. S., Hendrick, R. A., & Miller, M. F. (1981). The listener as a source of sound change. *parasession on language and behavior (Chicago Linguistics Society, Chicago 1981)*.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency. *Frequency and the emergence of linguistic structure*, *45*, 137-157.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Schertz, J., Kang, Y., & Han, S. (2019). Cross-language correspondences in the face of change: Phonetic independence versus convergence in two Korean-Mandarin bilingual communities. *International Journal of Bilingualism*, *23*(1), 157-199.

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

Winn, M. (2020). Praat script to manipulate VOT in natural speech. https://github.com/ListenLab/VOT.

Wright, J. D. (2007). *Laryngeal contrast in Seoul Korean*. University of Pennsylvania.