Age-indexed perceptual cue-shifting
of Korean stops in various prosodic positions[*]

Hyunjung So[**] · Jonny Jungyun Kim[***]
(Pusan National University)

So, Hyunjung and Jonny Jungyun Kim. 2024. Age-indexed perceptual cue-shifting of Korean stops in various prosodic positions. *Linguistic Research* 41(3): 367-390. Using an auditory identification task, this study examined the perceptual cue-weighting patterns of 29 younger Korean listeners as they distinguished between aspirated and lenis stops across various prosodic positions. The focus was on whether their perceptual routines reflected the ongoing sound change, namely the cross-generational shift from VOT to F0 in feature primacy. Consistent with the known prosodic constraint of this sound change, listeners relied more heavily on F0 cues in phrase-initial positions, while VOT retained feature primacy in phrase-medial contexts. Importantly, listeners adjusted their cue-weighting strategies depending on the talker's age. When listening to younger talkers, as compared to older talkers, listeners were more responsive to F0 cues and reduced their reliance on VOT cues. These age-related effects occurred only for phrase-initial stops. Our findings suggest that Korean stop categorization is flexibly informed by both talker characteristics and speech prosody based on listeners' prior experiences, which would facilitate cross-talker adaptation, and that the age-indexed perceptual flexibility is categorically governed by the higher-order intonational structure of Seoul Korean. Overall, the results contribute to the understanding of how social and linguistic factors interact in shaping mental representations of a sociophonetic variable under sound change. **(Pusan National University)**

Keywords Korean stops, sound change, speech perception, cue-weighting, prosody, talker adaptation, experience

---

1. Introduction

The typologically rare 3-way voiceless stop contrast in Korean has traditionally been described as a VOT-based system: (1) aspirated /pʰ, tʰ, kʰ/ with long lag VOT; (2) fortis /p*, t*, k*/ with short lag VOT; and (3) lenis /p, t, k/ falling in between but closer to the fortis group (Lisker and Abramson 1964). Over the past three decades, Korean stops—particularly the aspirated and lenis stops, which are the focus of this study—have undergone a sound change, as schematized in Figure 1.
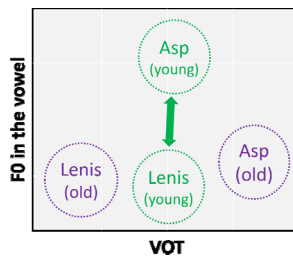

Figure 1. Cross-age difference in VOT and F0 cues for aspirated and lenis stops

The change seems to have been initiated by younger speakers in the 1990s, who began producing lenis stops with increased VOT in phrase-initial positions, leading to a subsequent overlap with aspirated stops rather than fortis stops (Silva 1992). As a result, aspirated and lenis stops produced by younger speakers are no longer distinguishable using the VOT cue phrase-initially, whereas older speakers still maintain the VOT-based distinction in all prosodic positions. This VOT merger has been consistently evidenced in a large body of empirical studies (see Lee et al. 2020 for a review), that examined speech production (Silva 2002, 2006; Wright 2007; Kang and Guion 2008; Oh 2011; Kim 2012, 2013, 2014; Kang 2014; Bang et al. 2018; Choi et al. 2020; Kim 2024), as well as perception (Lee and Jung 2000; Kim et al. 2002; Wright 2007; Kang 2010; Kong et al. 2011; Lee et al. 2013; Oh et al. 2018; Schertz et al. 2019). These studies also corroborated that the loss of VOT cues has been compensated by F0 cues, both acoustically and perceptually, such that aspirated stops (along with fortis stops) are associated with high F0 in the following vowel, and lenis stops with low F0.

This cue-shifting towards F0 as a primary cue is often described as being *in*

progress, rather than trans-phonologization of a new distinctive feature (e.g., Choi et al. 2020). This is not only because the innovative form remains age-indexed (i.e., associated with younger speakers), but also because it occurs in a prosodically limited context (i.e., only in prosodically strengthened, phrase-initial positions). To provide further evidence for the status of the ongoing sound change from speech perception, this study investigates how flexibly (if not robustly) the F0 feature is linked to younger generation's mental representations of the aspirated and lenis stops. To this end, we examine the extent to which younger listeners—who are presumably participating in (or at least have perceptually adapted to) the sound change—shift their cue-weighting pattern according to perceived talker age and varying prosodic positions.

2. Background and predictions

2.1 The social factor

Echoing with the sound change pattern described above, some perception studies observed that younger listeners rely more on F0 rather than VOT when distinguishing aspirated and lenis stops (Lee and Jung 2000; Kim et al. 2002; Wright 2007; Kang 2010). Kang (2010) directly compared the cue-weighting patterns used by younger listeners in their 20s (mean age = 20.2) versus older listeners over 40 (mean age = 48.5), and found that F0 was more highly weighted by the younger listener group than the older group. Similar results were reported in Wright (2007). Production data published around the same time (Kang and Guion 2008) also showed that F0 cues were enhanced in clear speech when compared to casual speech. While these findings shed light on the phonetic target of younger generations that has already shifted towards F0, evidence for individual listeners' *flexibility* in cue-shifting adaptations in accordance with the *talker age* is limited. Nevertheless, it is worth mentioning three experiments that examined socially-indexed cue-weighting strategies.

First, Schertz et al.'s (2019) experiment with bilingual Korean listeners in two different regions of China provides particularly valuable insights. They tested listeners in Hunchun, where the conservative VOT-based system was maintained, and in Dandung, where the innovative F0-based system had been adopted. The results indicated that cue-weighting patterns in identification responses for stop-initial target

syllables were significantly influenced by a set of listeners' social characteristics including age, region, and prior experience with dialects. Crucially, cue-weighting patterns were also affected by listeners' *expectations* about the two talkers used in the experiment, who differed in age. When listening to the younger female talker's voice, listeners relied primarily on F0 (rather than VOT), compared to when listening to the other talker who was older male. This effect was observed even among the conservative Hunchun listeners, though with a lesser extent than among the innovative Dandung listeners. Furthermore, the younger listener group showed a more pronounced use of F0 cues when informed that the talker spoke Seoul Dialect, rather than the listener's own dialect.

Second, using a different sociophonetic variable in Korean, Kang et al. (2024) tested cue-weighting patterns in the perception of two back vowels. In this sound change, younger Korean speakers, particularly females, raise /o/, which pushes /u/ forward to avoid overlap, resulting in F2 (rather than F1) becoming the primary acoustic cue to distinguish the two vowels. Their experiment played vowel sounds with varying F1 and F2 to 16 different types of talker-listener dyads, differing in both age and gender. Mirroring the reduced F1 cue weight in innovative speech, listeners in the most innovative dyad condition (i.e., younger female listeners hearing a younger female talker) primarily relied on F2 for identification, while F1 remained the primary cue in the most conservative dyad condition (i.e., older male listeners hearing an older male talker).

Last, to examine shifts in cue-weighting under ethnicity-indexed talker variability, Zhang and Kim's (2022) experiment manipulated the spectral quality (i.e., F1 and F2) and the length of an English tense-lax vowel pair (e.g., /i/ in *seat* or /ɪ/ in *sit*). They found that native English listeners significantly reduced their reliance on the spectral cues when the auditory target was preceded by a Chinese-accented carrier phrase. The inhibition of the primary cue in the listeners' native language was interpreted as stemming from their *previous experiences* with Chinese L2 speakers, whose English vowel tenseness contrast typically relied on vowel duration rather than spectral cues.

Experimental literature in sociophonetics has established the theoretical foundation for how experience-based expectations about the talker affect speech perception. For example, segment categorization is informed by social information attributed to the talker—such as age, gender, or region of origin, to name a few—whether conveyed

visually (Hay et al. 2006; Koops et al. 2008; Drager 2011), auditorily (Strand and Johnson 1996; Van Berkum et al. 2008), through verbal labels (Johnson et al. 1999; Niedzielski 1999), or even via unattended primes (Hay and Drager 2010). This body of literature, along with the talker-adaptation effects on cue-weighting patterns outlined above resonate with experience-based theories of speech processing, including exemplar-based models (Johnson 1997; Goldinger 1998; Pierrehumbert 2002; Hay and Foulkes 2016; Hay 2018; see also Frisch 2018 for a review). In this approach, phonetic memories are not stored in isolation but are encoded as phonetically rich clusters of episodic memories, contextually indexed with regularly observed linguistic or non-linguistic covariates, including talker characteristics.

In this regard, Schertz et al.'s (2019) results about the effect of talker age on Korean stop cue-weighting introduced above underscore the role of prior experiences as *listeners* of Seoul Korean in perceptual cue-shifting, regardless of an individual's production routines in relation to the sound change. Specifically, the bilingual population's phonetic representations for stop segments under sound change appeared to be multi-faceted, allowing for varying perceptual cue-weighting strategies consistent with previously accumulated correlation between social and phonetic information. This leads to the prediction that young speakers of Seoul Dialect residing in Korea would also exhibit such age-indexed perceptual shifts as those in China.

More precisely, they would show a general tendency of greater sensitivity to F0 differences while categorizing phrase-initial stops regardless of talker age, in line with previous findings (Lee and Jung 2000; Kim et al. 2002; Wright 2007; Kang 2010). At the same time, they would increase their reliance on F0 when listening to younger talkers, but shift to relying more on VOT when listening to older talkers. Alternatively, with more consistent daily exposure to, and more active use of, the F0-based contrast compared to the listeners in China, it is also plausible to assume that younger listeners in Korea possess stable long-term phonetic representations for F0-based distinction. If so, they may not shift between F0 and VOT according to talker age.

2.2 The prosodic factor

The main hypothesis above will be tested across various prosodic positions, since the sound change in question is structurally related to prosodic landmarks in Seoul Korean

intonation as introduced below (see Kang 2014; Cho and Lee 2016; Choi et al. 2020; Kim 2024, for a more detailed discussion).

The segment-dependent F0 cues (i.e., aspirated=high, lenis=low) interact closely with the Intonational Phonology of Seoul Korean proposed by Jun (1993, 1998, 2000). In this framework, the first two syllables in an Accentual Phrase (AP) or phrases at higher levels, such as an Intonational Phrase (IP), are assigned HH edge tones when the phrase-initial segment is [+ stiff vocal folds], or are assigned LH otherwise. This structural association was further elaborated by Kang (2014), who examined cross-generational corpus data of several types of phrase-initial consonants. This study demonstrated that consonants with [+ spread glottis] (e.g., /h/) induced an H tone while consonants without the feature (e.g., /n/) did not. This suggests that the enhanced F0 use of younger generation systematically targets the entire natural class of aspirated consonants, rather than the aspirated stops only. Therefore, as Choi et al. (2020) interpreted, in the presence of the phrase-initial tones assigned by higher-order prosodic-structure planning, VOT becomes redundant in terms of the stop phonation contrast, facilitating the effort-saving shift towards the innovative F0-based system.

In phrase-medial (i.e., word-initial) position, on the other hand, the tone is also assigned by the intonational phonology but not referring to the phonation-based phonemic category of the syllable-initial consonant. For this reason, VOT remains significantly different between aspirated and lenis stop categories regardless of speaker age, as demonstrated in production experiments that compared phrase-initial vs. phrase-medial positions in factorial design (Choi et al 2020; Kim 2024). Note that this phrase-medial VOT difference is also partially explained by the intervocalic voicing rule applied to a lenis stop within the AP domain (Jun 1998; Cho and Keating 2001).

Considering the prosodically restricted use of F0 cues for stop categories, our prediction above is expected to hold only in phrase-initial positions (i.e., IP-initial and AP-initial), while AP-medial stops (i.e., stops at the initial position of a prosodic word unit; henceforth, Wd-initial or phrase-medial stops) would not show individual listeners' cue-shifting. Interestingly, however, there is evidence that F0 difference is also substantially utilized phrase-medially in production, even in the presence of the primary role of VOT differences. Choi et al. (2020) showed that F0 was differentially used by both younger and older speaker groups, unless the phrase-medial target word was contextually unfocused. In addition, Kim's (2024) statistical comparison showed more pronounced phonation-dependent F0 separation for phrase-medial stops

produced by younger speakers.[1] Thus, we cannot entirely rule out the possibility of F0 serving as a primary cue in the perception of Wd-initial stops, either regardless of talker age or only in the younger-talker condition.

In sum, we examine how the two factors contributing to the realization of Korean stops compromise each other during speech perception. Using a 2-alternative forced choice auditory identification task, we test if perceptual cue-weighting pattern of VOT and F0 systematically varies depending on talker age, and also verify whether this age-indexed cue-shifting is limited in phrase-initial position, such as IP-initial and AP-initial.

3. Method

3.1 Participants

The experiment participants were 29 young college students (mean age=24, max=28, min=18; 18 females, 11 males), all of whom were born in the Seoul Metropolitan area and had lived in Seoul for more than 10 years. At the time of participation, 22 were residing in Seoul, and 7 in Busan. The experiment was conducted in each location. All participants, except one, identified as a 'Seoul person (서울 사람)' . Despite some sociolinguistic heterogeneity, they were all Seoul Dialect speakers with a Korean identity, unlike the participants in Schertz et al. (2019).

3.2 Talkers

Auditory targets were recorded by four talkers: OF (older female), OM (older male), YF (younger female), and YM (younger male). All of them were Seoul Dialect speakers having lived in Seoul for more than 20 years. Table 1 summarizes their age and perceived age. Perceived age was obtained from a post-experiment survey, during which participants were played one of the auditory stimuli per talker (VOT: 14ms,

---

1   This result, however, was observed in an experimental setting where the target stops were always preceded by another stop that matched in laryngeal feature (e.g., a phrase-medial aspirated stop following an utterance-initial aspirated stop). Since a high-tone-inducing segment can raise the overall pitch range of the phrase (Cho and Lee 2016), this result alone is not enough to support the F0-based distinction in phrase-medial stops.

F0: Step 2), and rated each talker's age by providing a specific number, such as 23 or 68, rather than an approximation, such as 'about 20' or 'late 60s'.

Table 1. A summary of talker age information

Talker ID	Age	Age perceived by participants			
		Mean	Min.	Max.	S.D.
OF	76	64.17	40	80	9.23
OM	60	56.21	40	73	7.38
YF	21	23.14	17	27	2.62
YM	25	28.03	20	40	4.04

3.3 Auditory stimuli

3.3.1 Baseline tokens

A stop-initial minimal pair of /toki/, 'ceramic', and /tʰoki/, 'pottery' was used as target words, based on the assumption that they are similar in meaning, with no critical difference in age-wise usage frequency. Although /toki/ may be somewhat less familiar to people in their twenties than /tʰoki/, we assumed that this difference was not substantial enough to introduce an age-related confounding effect arising from the words themselves (e.g., more VOT-based responses for a more old-associated word). To elicit natural realization of the targets in three different prosodic positions, a 43 year-old male native Seoul Dialect speaker trained in the K-ToBI system (the second author) pre-recorded six sample utterances, the pitch contours of which are shown in Figure 2. The four talkers were instructed to imitate the model talker's intonation as closely as possible while recording the auditory stimuli.
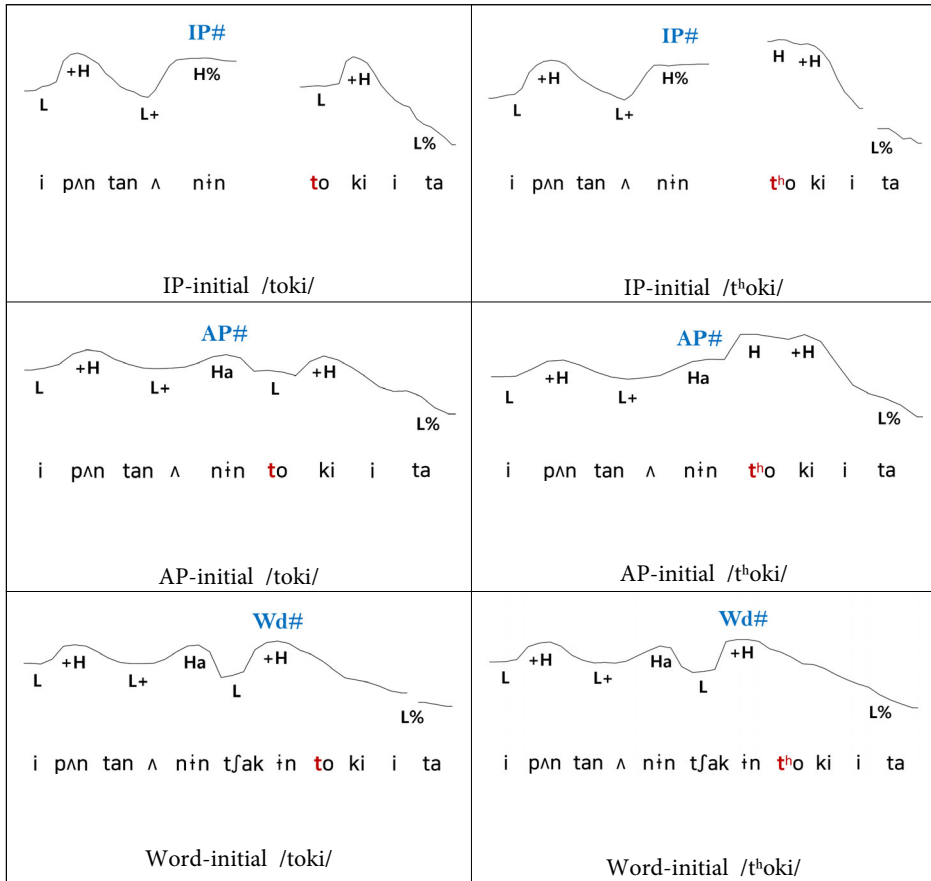
Figure 2. Pitch contours extracted from the model talker's utterances

Note that we used an identical sentence for both the IP-initial (top panels) and AP-initial (mid panels) conditions. In the IP-initial condition, however, the target word appeared after a pause, which forms an IP boundary along with the final lengthening of the carrier phrase /ipʌn tanʌnɨn/ (이번 단어는, 'the word is …'). In the AP-initial condition, the target occurred immediately after the right edge-tone marking in the first AP (i.e., L+Ha falling on /tanʌnɨn/), with no pause before the target.

While the target syllable received either L or H depending on the stop identity in these two phrase-initial conditions, targets in the Wd-initial condition (bottom

panels) occurred while F0 was falling from the left edge-tone marking in the second AP (i.e., L+H on the adjective /tʃakɨn/ (작은, 'small')). Note that although this target syllable was under influence of the interpolation from this edge tone (H) to the boundary tone (L%), the aspirated target /tʰo/ is slightly higher in pitch (mean F0 of the target syllable=124Hz) than the lenis target /to/ (mean=120Hz). We inserted the adjective /tʃakɨn/ (작은, 'small') ending with an [n] before the target word, forming a 6-syllable single AP, either /tʃakɨn toki-ita/ or /tʃakɨn tʰoki-ita/. This was done to make the target occur phrase-medially, while ensuring a homorganic environment (i.e., an alveolar stop preceded by an alveolar nasal) across all conditions.

During the recording session where the four talkers imitated the model utterances, care was taken for the AP-initial and Wd-initial conditions to avoid an IP boundary after /ipʌn tanʌnɨn/. For the Wd-initial condition, efforts were made to minimize the juncture between the adjective /tʃakɨn/ and the target, with the F0 value of /ɨn/ peaking and then immediately followed by an L%, showing continuous F0 falling until the end. Also, to prevent reduction, syllable-timed speech rhythm was maintained, making the tone differences across the syllables clear.

3.3.2 Acoustic manipulations

We manipulated the length of aspiration duration (VOT) and F0 using Praat PSOLA algorithm, referring to Schertz et al.'s (2019) method. A total of 360 stimuli were created with 6 VOT steps and 5 F0 steps for each of the 4 talkers and 3 prosodic positions, employing the matched-guise technique (Drager 2018; see Lambert et al. 1960, for the original design). That is, acoustic manipulation was applied only to the target syllable, which was then spliced into the complete utterance. This was done to isolate the effects of VOT and F0 manipulations from the effects of other factors, including individual voices and the length of closure duration before the release burst of Wd-initial targets.[2]

Regarding VOT manipulation, we further controlled for cross-talker variation in acoustic properties within the VOT duration. First, we extracted the release burst

---

2   As an anonymous reviewer pointed out, variation in the length of closure duration can confound the results in the Wd-initial condition. To address this, we ensured that all manipulated targets in the homorganic sequence /tʃakɨn tʰoki/ included a release burst occurring immediately after the voice bar in /n/, with no gap in vocal energy between them.

noise and aspiration period from an aspirated token recorded by OF. Second, using this token, we created 6 different VOT-step tokens for each prosodic position. Last, we spliced them with each talker's vowel onset from a lenis token. Through this procedure, VOT length systematically varied while the release burst noise was uniformly set to 2ms, with no cross-talker differences in breathiness and intensity within the aspiration duration.

Since VOT lengthens after a larger prosodic boundary (Cho and Keating 2001; Choi et al. 2020; Kim 2024), we adjusted both the interval between VOT steps and the overall VOT range according to boundary strength (IP > AP > Wd). As shown in Table 2, we applied the widest VOT range (2-82 ms) with a 16 ms interval for IP-initial stops, a range of 2-62 ms with a 12 ms interval for AP-initial stops, and a range of 2-42 ms with an 8 ms interval for Word-initial stops. These specific interval values (i.e., 16, 12, 8) were chosen to enhance naturalness in the manipulated outputs.

Table 2. The ranges and step intervals of VOT and F0 manipulation

OF	IP	AP	Wd	OM	IP	AP	Wd
VOT min (ms)	2	2	2	VOT min (ms)	2	2	2
VOT max (ms)	82	62	42	VOT max (ms)	82	62	42
VOT step interval (ms)	16	12	8	VOT step interval (ms)	16	12	8
F0 min (Hz)	164	164	185	F0 min (Hz)	92	97	125
F0 max (Hz)	244	244	245	F0 max (Hz)	172	165	185
F0 step interval (Hz)	20	20	15	F0 step interval (Hz)	20	17	15
YF	IP	AP	Wd	YM	IP	AP	Wd
VOT min (ms)	2	2	2	VOT min (ms)	2	2	2
VOT max (ms)	82	62	42	VOT max (ms)	82	62	42
VOT step interval (ms)	16	12	8	VOT step interval (ms)	16	12	8
F0 min (Hz)	153	168	204	F0 min (Hz)	82	88	101
F0 max (Hz)	233	248	264	F0 max (Hz)	142	148	161
F0 step interval (Hz)	20	20	15	F0 step interval (Hz)	15	15	15

For F0 manipulation, we considered each talker's F0 range along with the naturalness of the output. Due to the narrower F0 range of male talkers, OM's and particularly YM's F0 outputs had smaller ranges (and therefore smaller intervals) compared to those of OF and YF, as shown in Table 2.

In particular, for phrase-initial stops (IP-initial and AP-initial), all F0 values of the target syllable were controlled to be the same as or lower than those of the following syllable /ki/, in line with the HH (typically associated with /tʰ/) or LH (typically associated with /t/) phrase-initial tones in Seoul intonation. With this constraint, we first increased F0 of the syllable /ki/ to a level that sounded natural for each talker, then set talker-specific intervals for the F0 steps in the target syllable within a naturally sounding range. Since YM had a particularly low F0 range, the smallest F0 interval (15 Hz) was applied for both IP-initial and AP-initial positions.

Next, for Wd-initial stops, we adjusted the intonational pattern in the entire second AP (/tʃakɨn toki-ita/ or /tʃakɨn tʰoki-ita/) under two conditions. First, the target syllable was controlled to be equal to or lower than the preceding syllable /ɨn/ (as part of the adjective /tʃakɨn/) to avoid marking the target syllable as the start of a new, separate AP. To secure a sufficient range for manipulating the F0 of the target syllable, the F0 of the preceding syllable /ɨn/ was raised within a natural range. Second, F0 of the utterance-final syllable /ta/ (as part of the sentence-ending marker /-ita/) was fixed per talker, and the F0 of the target syllable was manipulated to always be higher than the final /ta/, thereby naturally inducing an L% boundary tone to signify a declarative sentence. Depending on the F0 step of the target syllable, F0 slopes throughout the /toki-i/ or /tʰoki-i/ portion were adjusted to sound as natural as possible, with a linear F0 drop towards the final syllable /ta/.

3.4 Procedure

The experiment was conducted using E-prime software on a laptop in a quiet room. Participants were instructed to distinguish whether the sound they heard was /toki/ (lenis-initial) or /tʰoki/ (aspirated-initial) across 360 trials, by pressing the Caps Lock key for lenis or the Enter key for aspirated responses. All stimuli were presented in a random order per participant. Four practice trials were presented to help familiarize themselves with the task.[3] To minimize the potential influence of interaction

with the experimenter, who is a younger female Kyeongsang Dialect speaker, detailed instructions were displayed on the monitor for participants to read before the experiment. After the experiment, background information of the participants was collected via a questionnaire.

3.5 Analysis

Among the 10,440 responses collected (360 trials × 29 participants), 21 tokens were excluded because the key press was recorded by E-Prime before the stop consonant was heard in its entirety. To statistically test the remaining data, mixed-effects logistic regression models were constructed for each position (IP, AP, Wd) using the lmerTest package in R 4.4.1. We report results from these three separate models, rather than overall results pulled across all three conditions, since our hypotheses are related to the significance of the talker age effect on the cue-weighting patterns in each prosodic position.[4]

The binary dependent variable for all models was Response (lenis, aspirated), with the reference level indicated by underlining. For fixed effects, we first included the main effects of VOT (steps 1 to 6) and F0 (steps 1 to 5), both treated as continuous variables, as well as two critical interaction terms: (1) the interaction between VOT and Talker Age (old, young), and (2) the interaction between F0 and Talker Age.[5] Next, through likelihood ratio tests (Winter 2020), Talker Gender (female, male) was added as a control factor.[6] The two binary variables, Talker Age and Talker Gender were sum-coded as –0.5 and 0.5.

Additionally, we tested main effects of Talker Age, Participant Gender, Trial, Word Duration, Relative VOT (raw VOT divided by Word Duration), Location (Seoul, Busan), Participant Identity (Seoul, others), Perceived Talker Dialect (Seoul, others), and 3-way interactions between each of these factors and the two critical interaction

---

3   To avoid prior exposure to the voice used in the experiment, voices generated by text-to-speech (TTS) technology were used.

4   Even if a global analysis shows a significant interaction between the talker age effect and prosodic position (e.g., a greater effect magnitude in phrase-initial than in phrase-medial), it does not provide evidence for the significance of the age effect in each position.

5   While these interaction terms are directly related to our research questions, the main effect of Talker Age was excluded to avoid over-fitting.

6   Although Talker Gender did not significantly improve all models' fit, it was uniformly included in all three final models for the sake of fair cross-model comparison.

terms. However, all of these were excluded as they did not reach significance consistently across the models. Random effects were structured minimally, including only intercepts for both item and subject, as random slopes failed to converge in all models.

4. Results

4.1 IP-initial position

We first examine the response patterns to IP-initial stops. As visualized in Figure 3 (a) and (b), an increase in either VOT or F0 steps on the horizontal axis respectively induced a higher rate of responses for the aspirated stop /t^h/. Accordingly, the main effects of VOT and F0 were significant ($p<.001$ for both), as shown by the model's predicted values summarized in Table 3. While this indicates that our younger listeners were generally sensitive to the change in both phonetic variables, a comparison of the effect magnitudes based on the estimated β-slope values reveals that they were more reactive to the F0 increase (β=1.235) than the VOT increase (β=0.795).
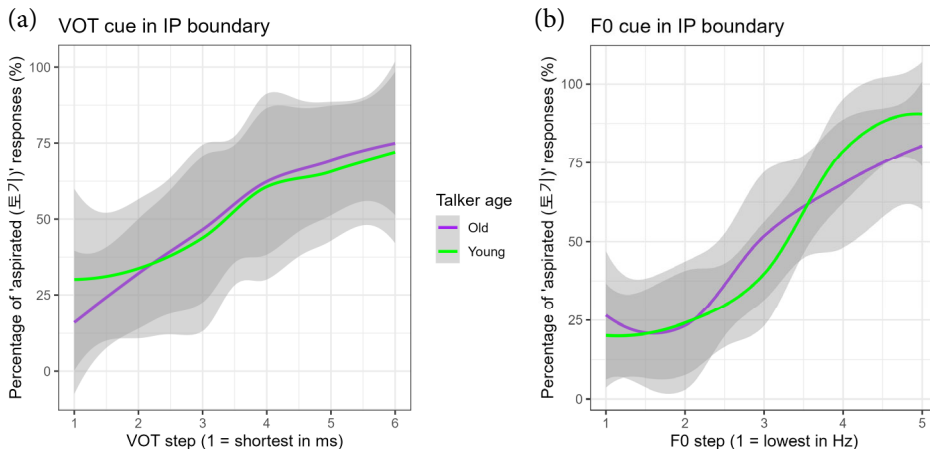

Figure 3. Percentage of aspirated responses for IP-initial stops by (a) VOT and (b) F0 steps, and their interaction with talker age: Covariates of the regression lines are estimated using the loess smooth method. Grey areas indicate standard errors.

Table 3. Results of the mixed-effects logistic model for IP-initial stops

IP-initial model	Estimate	Std. Error	z-value	p-value
(Intercept)	-6.442	0.375	-17.156	<.001
VOT	0.795	0.057	13.967	<.001
F0	1.235	0.071	17.473	<.001
Talker Gender = Male	-0.047	0.180	-0.264	0.792
VOT : Talker Age = Young	-0.174	0.074	-2.354	0.019
F0 : Talker Age = Young	0.258	0.089	2.901	0.004

Importantly, the VOT : Talker Age interaction showed that the effect of VOT was significantly reduced when listening to younger talkers compared to older talkers ($p<.05$). The F0 : Talker Age interaction also showed that listeners' sensitivity to F0 increased when listening to younger talkers compared to older talkers ($p<.01$). Taken together, these results indicate an effect of talker age on cue-weighting between VOT and F0. Based on the β-slope coefficients, the effect of younger talkers appeared to be primarily driven by strengthening of F0-based perceptual categorization (β=0.258), which compensated the reduced sensitivity to VOT-based categorization (β=-0.174).

4.2 AP-initial position

Results from AP-initial stops are shown in Figure 4 and Table 4. In line with the IP-initial results, the main effects of VOT ($p<.001$) and F0 ($p<.001$) indicate listeners' general sensitivity to both, and the effect magnitude was greater for the change in F0 (β=1.349) than in VOT (β=0.712). Additionally, Talker Age interacted both with VOT ($p<.05$) and with F0 ($p<.001$), suggesting greater reliance on F0 and reduced reliance on VOT while listening to younger talkers' AP-initial stops. This interaction was also primarily led by strengthening of the F0 cue (β=0.332), rather than inhibiting the VOT cue (β=-0.193).
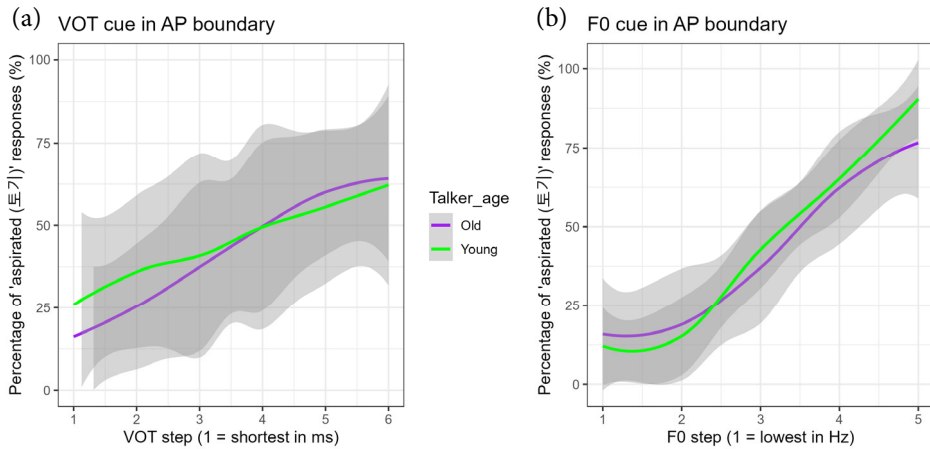
Figure 4. Percentage of aspirated responses for AP-initial stops

Table 4. Results of the mixed-effects logistic model for AP-initial stops

AP-initial model	Estimate	Std. Error	z-value	p-value
(Intercept)	-7.026	0.408	-17.240	<.001
VOT	0.712	0.060	11.800	<.001
F0	1.349	0.076	17.725	<.001
Talker Gender = Male	1.077	0.195	5.534	<.001
VOT : Talker Age = Young	-0.193	0.080	-2.410	0.016
F0 : Talker age = Young	0.332	0.096	3.473	<.001

Unlike the IP model, however, the AP model showed a main effect of Talker Gender (p<.001), suggesting that participants were more likely to perceive the sound as the aspirated /tʰ/ when produced by male talkers (whether young or old) compared to female talkers. While we cannot provide a compelling explanation for why this gender effect occurred, further analysis confirmed that it was primarily driven by a low rate of aspirated responses when listening to the two female talkers, particularly in the AP-initial condition: OF (overall=48%, AP=39%); YF (42%, 34%). Male talkers' aspirated response rates were closer to the chance level (50%): OM (49%, 46%); YM (53%, 57%). Thus, it is inferred that the female voices introduced a bias away from aspirated responses.[7]

4.3 Wd-initial position

As shown in Figure 5 and Table 5, Wd-initial stops also tended to be categorized as an aspirated /t^h/, as VOT or F0 increased ($p<.001$ for both effects). Unlike previous phrase-initial conditions, however, our young listeners weighted the VOT cues (β =1.082) above the F0 cues ($\beta=0.955$), indicating the feature primacy of VOT in the perception of phrase-medial stops. With the reduced role of F0 cues in identifying Wd-initial stops,[8] age-indexed cue-shifting did not arise: Talker Age did not interact with either VOT ($p=.440$) or F0 ($p=.061$).

---

7   As an anonymous reviewer suggested, the bias may be due to incomplete controls for acoustic cues beyond VOT and F0. Alternatively, we suspect that it arose from the F0 manipulation method. Recall our method for IP-initial and AP-initial stops in Section 3.3.2: the target syllable was never higher than the following syllable /ki/. While this restriction aligns with the two possible phrase-initial edge tones in Seoul intonation (HH and LH), it is, in fact, also possible for the target syllable to be higher than /ki/ and still be labeled as HH. We speculate that excluding this possibility in our stimuli resulted in the generally low probability of aspirated responses because even the second highest F0 step could potentially be perceived as LH (rather than HH), biasing towards a lenis response. This bias appeared stronger for the female talkers who utilized a wider pitch intervals than the males. Moreover, this factor most critically affected AP-initial stops, in which reduced F0 intervals were used for both male talkers (by 3Hz for OM, by 5Hz for YM; see Table 2). The rationale above also explains why the YM talker, whose F0 was also reduced by 5Hz for IP-initial stops, induced the highest aspirated response rate in general (overall=53%, AP=57%).

8   From a more conservative viewpoint, the significant effect of F0 on perception here may be interpreted as mirroring the low-level F0 perturbation effect in production. Since aspirated stops are associated with greater laryngeal tension, which in turn increases F0 in vowel onset intrinsically, it is more likely for the listeners to hear an aspirated stop for high-pitched syllables.
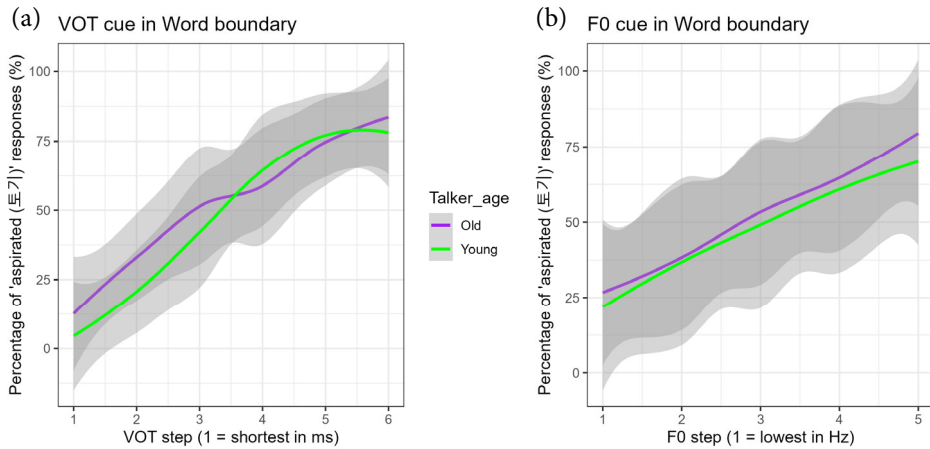
(a) VOT cue in Word boundary

(b) F0 cue in Word boundary

Figure 5. Percentage of aspirated responses for Wd-initial stops

Table 5. Results of the mixed-effects logistic model for Wd-initial stops

Wd-initial model	Estimate	Std. Error	z-value	p-value
(Intercept)	-6.660	0.364	-18.293	<.001
VOT	1.082	0.057	18.996	<.001
F0	0.955	0.065	14.637	<.001
Talker Gender = Male	0.275	0.171	1.614	0.107
VOT : Talker Age = Young	0.054	0.070	0.772	0.440
F0 : Talker Age = Young	-0.156	0.084	-1.870	0.061

5. General discussion

The results above showed that young listeners' stop identification significantly depended on the changes in either F0 or VOT, ignoring the effects of varying talker age and prosodic positions. At the same time, F0 differences in general played a more pivotal role than VOT differences when identifying phrase-initial stops—as measured by the slope coefficients—but not for phrase-medial stops. This result is consistent with the previous evidence (Lee and Jung 2000; Kim et al. 2002; Wright 2007; Kang 2010) showing that younger generations' perceptual target has been calibrated in line

with their own speech style, which features F0-based stop distinction in phrase-initial position. In other words, perceptual representations of Korean stops appear to have been shaped closely in line with the age-indexed production pattern summarized at the outset of this paper.

The primary result in this study further elaborated that younger listeners did not invariably rely on F0 cues, but they probabilistically adjusted the degree of their F0-reliance, showing an experience-based fine-tuning strategy referring to social and prosodic factors underlying the sound change. Specifically, under the younger-talker condition, listeners enhanced their preference for F0 cues in phrase-initial stops, while inhibiting the use of VOT cues. Taken together with Schertz et al.'s (2019) results sampled from listeners of Korean living in China, our data collected from those living in Korea (either in Seoul or Busan) provides evidence that experience-based speech processing is utilized to shift perceptual cue-weighting patterns, consistent with previous studies that used different social and phonetic variables (Zhang and Kim 2022; Kang et al. 2024).

The mental processes of cue-shifting for phrase-initial stops can be accounted for in exemplar-based terms (Johnson 1997; Goldinger 1998; Pierrehumbert 2002; Hay and Foulkes 2016). Since the age cues were extractable solely from the auditory input, as soon as the initial acoustic input (i.e., the carrier phrase) was retrieved, listeners would formulate implicit expectations about the fine-grained phonetic properties for them to concentrate. At first, each talker's voice quality and/or pronunciation styles would activate age-consistent exemplars of prior experiences with phrase-initial stops. This was possible since these age-indexed phonetic exemplars had been clustered distinctively around the typicality of the two prominent phonetic cues (F0 or VOT). Thus, activation immediately spread to prime listeners' attention to the age-congruent phonetic cue. In this regard, perceptual representations for the phonological units undergoing a sound change are being flexibly shaped based on prior experiences of encountering fine-grained phonetic differences closely linked to social characteristics of the speakers in the speech community, leaving room for facilitated cross-talker adaptations.

To focus on the current status of the sound change, the background for our null-hypothesis needs to be revisited. All our participants were born in or near Seoul after 1995, and even their parents may have actively participated in the sound change. Throughout the lifetime, they have been exposed to the rise of F0 as a statistically

meaningful cue in all prosodic positions, unless reduced under no focus (Choi et al. 2020; Kim 2024). Further, considering that this sound change can be seen as the initial stage of tonogenesis in Korean phonology as discussed in Kang (2014) a decade ago, we were led to investigate the extent to which younger listeners currently living in Korea have accumulated *categorically separable* exemplar distributions of F0-indexed phrase-initial stops. If the tonogenetic process had progressed further, the F0 cues may have been used consistently for phrase-initial stops, not being affected by talker age, or even by prosodic position, during similarity-based perceptual mapping of the incoming signal with the long-term exemplar clusters.

However, this was clearly not the case. Although F0 was a statistically significant perceptual cue in all prosodic positions, its distinctive role as a phonetic cue was limited in phrase-initial positions. In addition, young listeners' perceptual cue-shifting patterns were also socially conditioned and prosodically limited, in close alignment with the ongoing sound change pattern observed in previous studies. In this regard, the age-indexed perceptual flexibility seems to be available only in phrase-initial position, being strictly governed by the prosodic structure of Seoul Korean, even though F0 cues may actually be as prominent as the VOT cues in the production of Wd-initial stops (Choi et al. 2020; Kim 2024).

Therefore, as a final remark, even if the F0-based stop distinction is taking place as part of tonogenesis, it must be currently at a very early stage, still being conditioned by social and prosodic information. More precisely, younger generations' perceptual representations for phrase-initial stops appear to be socially-indexed; however, such social indices are not necessary for phrase-medial stops.

6. Conclusion

In the context of the ongoing sound change of Korean stops, this study examined whether and how younger native Korean listeners' perceptual cue-weighting patterns for the stop sounds are strategically fine-tuned as a function of social and prosodic factors. The results confirm that while younger listeners showed a greater reliance on F0 cues in phrase-initial positions, VOT remained a primary cue for categorizing phrase-medial stops, indicating that young listeners' perceptual target has shifted hand-in-hand with their own production patterns. The primary results showed that

younger listeners' reliance on F0 increased (while that on VOT decreased) when listening to younger talkers, highlighting the socially-indexed flexibility of the perceptual mechanism of cross-talker adaptation. However, since this effect was limited in phrase-initial stops, the flexibility appears to be categorically governed by higher-ordered intonational structure of Seoul Korean. These findings also suggest that the ongoing shift from VOT to F0 as a primary cue in Korean stop perception is not a fully phonologized change but is influenced by both social and prosodic factors.

References

Bang, Hye-Young, Morgan Sonderegger, Yoonjung Kang, Meghan Clayards, and Tae-jin Yoon. 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. https://doi.org/10.1016/j.wocn.2017.09.005.

Cho, Sunghye and Yong-Cheol Lee. 2016. The effect of the consonant-induced pitch on Seoul Korean intonation. *Linguistic Research* 33(2): 299-317.

Cho, Taehong and Patricia A. Keating. 2001. Articulatory and acoustic studies on domain-initial strengthening in Korean. *Journal of Phonetics* 29(2): 155-190.

Choi, Jiyoun, Sahyang Kim, and Taehong Cho. 2020. An apparent-time study of an ongoing sound change in Seoul Korean: A prosodic account. *PLoS ONE* 15(10): e0240682. https://doi.org/10.1371/journal.pone.0240682.

Drager, Katie. 2011. Speaker age and vowel perception. *Language and Speech* 54(1): 99-121.

Drager, Katie. 2018. *Experimental research methods in sociolinguistics*. New York, NY: Bloomsbury.

Frisch, Stefan A. 2018. Exemplar theories in phonology. In Stephen J. Hannahs and Anna Bosch (eds.), *The Routledge handbook of phonological theory*, 553-568. New York, NY: Routledge.

Goldinger, Stephen D. 1998. Echoes of echoes? An episodic theory of lexical access. *Psychological Review* 105(2): 251-279.

Hay, Jennifer. 2018. Sociophonetics: The role of words, the role of context, and the role of words in context. *Topics in Cognitive Science* 10(4): 696-706.

Hay, Jennifer and Katie Drager. 2010. Stuffed toys and speech perception. *Linguistics* 48(4): 865-892.

Hay, Jennifer and Paul Foulkes. 2016. The evolution of medial /t/ over real and remembered time. *Language* 96(2): 298-330.

Hay, Jennifer, Paul Warren, and Katie Drager. 2006. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics* 34(4): 458-484.

Johnson, Keith. 1997. Speech perception without speaker normalization: An exemplar model. In Kieth Johnson and John W. Mullennix (eds.), *Talker variability in speech processing*, 145-165. San Diego, CA: Academic Press.

Johnson, Keith, Elizabeth A. Strand, and Mariapaola D'ImperioK. 1999. Auditory-visual integration of talker gender in vowel perception. *Journal of Phonetics* 27(4): 359-384.

Jun, Sun-Ah. 1993. *The phonetics and phonology of Korean prosody.* PhD Dissertation. Ohio State University.

Jun, Sun-Ah. 1998. The accentual phrase in the Korean prosodic hierarchy. *Phonology* 15(2): 189-226.

Jun, Sun-Ah. 2000. K-ToBI (Korean ToBI) labelling conventions. *Speech Sciences* 7(1): 143-170.

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

Kang, Kyoung-Ho and Susan G. Guion. 2008. Clear speech production of Korean stops: Changing phonetic targets and enhancement strategies. *The Journal of the Acoustical Society of America* 124(6): 3909.

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

Kang, Yoonjung, Suyeon Yun, and Na-young Ryu. 2024. Talker- and listener-conditioned use of height-dependent vowel duration cue under sound change in progress: /o/ to /u/ raising in Daejeon Korean. In Taehong Cho, Sahyang Kim, Jeff Holiday, and Sang-Im Lee-Kim (eds.), *Proceedings of the 19th Conference on Laboratory Phonology*, 243-244. Seoul: Hanyang University Press.

Kim, Jonny Jungyun. 2024. Prosodically conditioned VOT merger of Korean stops examined in verse-style broad-focus speech. *Journal of Language Science* 31(1): 111-134.

Kim, Mi-Ryoung. 2012. Tonogenesis in Korean: Some recent speculations on the sound change. *Korea Journal of Linguistics* 37(2): 243-283.

Kim, Mi-Ryoung. 2013. VOT merger and F0 maximization between the lax and aspirated stop in sound change. *The Linguistic Association of Korea Journal* 21(2): 1-20.

Kim, Mi-Ryoung. 2014. Ongoing sound change in the stop system of Korean: A three-to two-way categorization. *Studies in Phonetics, Phonology and Morphology* 20(1): 51-82.

Kim, Mi-Ryoung, Patrice Speeter Beddor, and Julie Horrocks. 2002. The contribution of consonantal and vocalic information to the perception of Korean initial stops. *Journal of Phonetics* 30(1): 77-100.

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

Koops, Christian, Elizabeth Gentry, and Andrew Pantos. 2008. The effect of perceived speaker age on the perception of PIN and PEN vowels in Houston, Texas. *University of Pennsylvania Working Papers in Linguistics* 14(2): 12.

Lambert, Wallace E., Richard C. Hodgson, Robert C. Gardner, and Samuel Fillenbaum. 1960.

Evaluational reactions to spoken languages. *The Journal of Abnormal and Social Psychology* 60(1): 44.

Lee, Hyunjung, Stephen Politzer-Ahles, and Allard Jongman. 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.

Lee, Hyunjung, Jeffrey J. Holliday, and Eun Jong Kong. 2020. Diachronic change and synchronic variation in the Korean stop laryngeal contrast. *Language and Linguistics Compass* 14(7): e12374.

Lee, Kyung-Hee and Myung-Sook Jung. 2000. Acoustic characteristics and perceptual cues for Korean stops. *Speech Sciences* 7(2): 139-155.

Lisker, Leigh and Arthur S. Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20(3): 384-422.

Niedzielski, Nancy. 1999. The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology* 18(1): 62-85.

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

Oh, Eunjin, Kaori Idemaru, and Boram Kim. 2018. The use of a voice onset time cue in the perception of Seoul Korean stops as a function of listener gender. *The Korean Journal of Linguistics* 43(4): 761-780.

Pierrehumbert, Janet. 2002. Word-specific phonetics. In Carlos Gussenhoven and Natasha Warner (eds.), *Laboratory phonology 7*, 101-140. Berlin; New York: De Gruyter Mouton.

Schertz, Jessamyn, Yoonjung Kang, and Sungwoo Han. 2019. Sources of variability in phonetic perception: The joint influence of listener and talker characteristics on perception of the Korean stop contrast. *Laboratory Phonology* 10(1): 1-32.

Silva, David James. 1992. *The phonetics and phonology of stop lenition in Korean*. PhD Dissertation. Cornell University.

Silva, David James. 2002. Consonant aspiration in Korean : A retrospective. In Sang-Oak Lee and Gregory K. Iverson (eds.), *Pathways into Korean language and culture: Essays in honor of Young-Key Kim-Renaud*, 447-469. Seoul: Pagijong Press.

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

Strand, Elizabeth A. and Keith Johnson. 1996. Gradient and visual speaker normalization in the perception of fricatives. In Dafydd Gibbon (ed.), *Natural language processing and speech technology: Results of the 3rd KONVENS conference*, 14-26. Berlin; Boston: De Gruyter Mouton.

Van Berkum, Jos J. A., Danielle van den Brink, Cathelijne M. J. Y. Tesink, Miriam Kos, and Peter Hagoort. 2008. The neural integration of speaker and message. *Journal of Cognitive Neuroscience* 20(4): 580-591.

Winter, Bodo. 2020. *Statistics for linguists: An introduction using R*. New York, NY: Routledge.

Wright, Jonathan D. 2007. *Laryngeal contrast in Seoul Korean.* PhD Dissertation. University of Pennsylvania.

Zhang, Yinjun and Jonny Jungyun Kim. 2024. Can talker accent change cue-weighting patterns in English vowel categorization? Presented at *the 18th Conference on Laboratory Phonology.* 25 June. Online.

Hyunjung So
Master's Student
Department of English Language and Literature
Pusan National University
2, Busandaehak-ro 63beon-gil, Geumjeong-gu,
Busan, 46241 Korea
E-mail: minny980@pusan.ac.kr

Jonny Jungyun Kim
Associate Professor
Department of English Language and Literature
Pusan National University
2, Busandaehak-ro 63beon-gil, Geumjeong-gu,
Busan, 46241 Korea
E-mail: jonnykkim@gmail.com