

JOURNAL ARTICLE

Sources of variability in phonetic perception: The joint influence of listener and talker characteristics on perception of the Korean stop contrast

Jessamyn Schertz^{1,2}, Yoonjung Kang^{2,3} and Sungwoo Han⁴

¹ Department of Language Studies, University of Toronto Mississauga, Mississauga, CA

² Department of Linguistics, University of Toronto, Toronto, CA

³ Centre for French and Linguistics, University of Toronto Scarborough, Toronto, CA

⁴ Department of Korean Language and Literature, Inha University, Incheon, KR

Corresponding author: Jessamyn Schertz (jessamyn.schertz@utoronto.ca)

Where there is dialectal variability in production of a sound contrast, listeners from the two dialects may show parallel differences in perception. At the same time, perception is not static and can be influenced by other factors, including listeners' experience with, and expectations about, different talkers. This work examines perception of the Korean three-way stop phonation contrast by listeners of two dialects of Korean. We examine to what extent listeners' perception reflects production norms in their local community and, via a reverse matched-guise task, test whether their knowledge of cross-dialectal variability plays an active role in the way they categorize the contrast. While perception appears to reflect production norms on a broad level, we found age-related differences in perception, even for listener groups who showed no sign of a parallel difference in production. Furthermore, listeners showed different response patterns depending on the apparent dialect of the talker. Our results suggest that exposure to dialectal variability and expectations about the talker influence perception.

Keywords: sociophonetics; Korean; dialectal variation; perception; listener expectations

1. Introduction

Dialectal differences in the realization of a phonetic contrast are generally expected to be mirrored in perception, presumably due to different phonetic representations that form the basis of both perception and production. At the same time, perception can be influenced by listeners' expectations about what they are likely to hear: A growing body of work has provided evidence that social information about the talker can independently influence perception (e.g., Niedzielski, 1999; Strand & Johnson, 1996; Hay, Warren & Drager, 2006b), although findings are complex and sometimes inconsistent (e.g., Squires, 2013; Lawrence, 2015). The current work examines the joint influence of characteristics of the listener and expectations about the talker in perception of the Korean three-way stop contrast by listeners of different dialects and ages. We track how native Korean listeners from Hunchun and Dandong, China use various phonetic cues to the contrast (VOT, f_0 , and voice quality of the following vowel) when listening to talkers said to be from different dialects. We test the extent to which regional and age-based differences in production norms account for differences in perceptual patterns, and explore how listeners modify their categorization of the contrast based on the apparent dialect of the talker.

1.1. Influence of listener and talker characteristics on speech perception

Under a traditional prototype-based view of phonetic representations, dialect-based variability in perception stems from the same source as the variability in production: dialect-specific phonetic prototypes. On a broad level, group differences in perception of phonetic contrasts have been found to reflect those in production. For example, Lee, Politzer-Ahles, and Jongman (2013) found that Kyungsang Korean speakers, who use f_0 less than Seoul speakers to distinguish lenis and aspirated stops in production, also rely less on f_0 than Seoul listeners in perception. Similarly, work on dialectal differences in vowel perception (e.g., Willis, 1972; Fridland & Kendall, 2012) has shown support for a relationship between overall production and perception patterns (see Thomas, 2002 for a review). However, close examination of these studies reveals that almost none have found a straightforward link between perception and production on all of the predicted contrasts: For example, out of the four predicted differences in Willis (1972), two were supported and two showed “problematic results” (Willis 1972, p. 246).

At the same time, a large body of recent work has explored how social knowledge and expectations influence speech perception. One of the first studies demonstrating the importance of expectations on perception was by Strand and Johnson (1996), who asked listeners to categorize identical /s/-/ʃ/ continua paired with videos of either a male or a female talker. Listeners showed different category boundaries depending on the talker shown in the video, requiring a higher frequency threshold for /s/ classification when the face shown in the video was female, reflecting the fact that the /s/-/ʃ/ boundary tends to be higher in frequency for females than for males. Subsequent work has provided evidence that expectations based on other social factors, including race (Staum Casasanto, 2010), social class (Hay, Nolan & Drager, 2006a; Dufour, Kriegel, Alleesaib, & Nguyen, 2014), persona (e.g. ‘Valley Girl,’ D’Onofrio, 2015), and talker age (Hay et al., 2006b; Koops, Gentry, & Pantos, 2008) also influence speech perception (see Drager, 2010 for a review). However, effects found in these studies are often very small and/or inconsistent, and null results have been found in replications (e.g., Lawrence, 2015; Pharaoh, Lundholm Appel, Wolter, & Thogersen, 2015; Chang, 2017). This suggests that although there is robust evidence that social information *can* have an effect on perception, there is much work to be done on understanding how, when, and why this influence occurs.

Few studies have looked at the interaction between listener- and talker-level influences. One notable exception, Evans and Iverson (2004), examined perception of the vowel in English words like ‘bud’ by listeners of two accents of British English that differ in their pronunciation of this vowel, which is produced with a lower F1 in Northern (Sheffield) English than in Southern Standard British English. Listeners heard target words differing in the formant frequencies of the vowel embedded in carrier phrases spoken in the two different accents. Results showed that both listeners’ regional affiliation and the talker’s accent affected listeners’ ratings. Northern listeners’ highest-rated tokens had lower F1 than those of Southern listeners. However, the accent of the carrier phrase also influenced responses in the expected direction, with both listeners preferring tokens with lower F1 when preceded by a Northern accent. These results indicate that both listener and talker information can jointly influence perception, such that neither plays a deterministic role.

Communities in which there is a sound change in progress provide a particularly rich context for exploring the influence of expectations on speech perception. Hay et al. (2006b) found support for their prediction that listeners would use their knowledge of an ongoing merger between the diphthongs in ‘near’ /iə/ and ‘square’ /eə/ in New Zealand English to inform phonetic categorization: Listeners showed less accurate identification of the diphthongs in younger speakers. Similarly, in the context of an ongoing ‘pin’-‘pen’ ‘unmerger’ in Houston, Texas, Koops et al. (2008) found that listeners were more likely

to assume a merged system when listening to an older (versus middle-aged) talker. These results indicate that listeners are sensitive to the social distribution of pronunciation variants and that they use this knowledge in speech perception.

While most such studies have focused solely on the age of the talker in predicting how social information will be used, Drager (2011) also examined how the influence of talker age might interact with the age of the *listener*. Drager found that older, but not younger, participants showed different vowel category boundaries depending on the apparent age of the talker. In other words, older listeners appeared to use age-related variability to inform their speech perception, while younger listeners did not. Drager hypothesized that older listeners, having witnessed the full time course of the sound change, might be more likely to be aware of age-based social patterning. We explore whether there are similar age-based differences in sensitivity to talker-level information in the current work.

1.2. Linguistic background

1.2.1. Korean three-way stop contrast: Variation and change

This work focuses on listeners' categorization of Korean stops across three acoustic dimensions: VOT, f_0 , and voice quality of the following vowel. The three-way Korean stop contrast (aspirated versus lenis versus fortis; e.g., /p^h/ versus /p/ versus /p'/) has traditionally been described, primarily for Seoul Korean, as differing in VOT (aspirated > lenis > fortis), f_0 (aspirated and fortis having higher f_0 than lenis). Voice quality of the following vowel also plays a role: Fortis stops are associated with a more pressed or creaky quality than the other two categories, and this has been quantified in previous work by H1-H2, the amplitude difference between the first two harmonics (e.g., Cho, Jun, & Ladefoged, 2002; Kim, Beddor, & Horrocks, 2002). These differences in f_0 and voice quality are largest at vowel onset but extend through the vowel midpoint (Cho et al., 2002).¹ The stop system has been undergoing a well-documented diachronic change in several dialects, including standard Seoul Korean: The VOT difference between lenis and aspirated stops has been decreasing in younger Seoul speakers (e.g., Silva, 2006), a change which is accompanied by concurrent enhancement of the f_0 distinction in younger speakers' productions (Kang, 2014).

However, not all dialects are undergoing this change. While a Seoul-like pattern of change in VOT use has been found in Korean speakers from Shenyang, China (Jin, 2008) and Jeju (Holliday & Kong, 2011), a large VOT difference between lenis and aspirated stops is maintained by younger speakers of Kyungsang Korean (Lee & Jongman, 2012; Jang, 2012), although more recent work on Kyungsang does report an age-related change parallel, albeit at an earlier stage, to that found in Seoul (Lee & Jongman, 2018). No evidence of a change was found in Yanbian Korean, spoken in China (Oh & Yang, 2013). As noted by these authors, a plausible explanation for this apparent resistance to a VOT merger is that both the Kyungsang and Yanbian dialects have lexical pitch accent, which was present in Middle Korean but subsequently lost in Western dialects. Those dialects in which a change has been reported, including Seoul, all lack lexical pitch accent, whereas those in which there is less or later change are those which maintain the lexical pitch accent contrast. In these dialects, since f_0 is already used for a lexical distinction, it may be more difficult for f_0 to 'take over' as a phonetic cue to the laryngeal distinction as well, and this may in turn inhibit any potential VOT merger.

¹ Voice quality is a complex phenomenon, and there are many acoustic and articulatory correlates (e.g., Gordon & Ladefoged, 2001; Hillenbrand, Cleveland, & Erickson, 1994). While H1-H2 has been overwhelmingly used as a measure of creakiness/breathiness in previous literature on Korean stops, Cho et al. (2002) also took another spectral slope measure: H1-F2 (the difference between the amplitude of the first harmonic and the second formant), a proposed correlate to abruptness of vocal fold closure, and found similar voice quality differences at both onset and midpoint.

Dialectal differences in the relative importance of VOT and f_0 in distinguishing the lenis-aspirated contrast have been shown to be mirrored in perception: For example, Kyungsang Korean listeners rely more heavily on VOT, and less on f_0 , when distinguishing the lenis-aspirated contrast in comparison with Seoul listeners (Jang, 2012; Lee et al., 2013). There has been little work on dialectal differences in the perception of cues other than f_0 and VOT, but Ito and Kenstowicz (2008) propose that voice quality of the following vowel constitutes the primary perceptual cue for distinguishing fortis from lenis stops in Yanbian Korean (a dialect related to the one spoken in Hunchun), and support this with judgments from a single Yanbian listener on stimuli that had been cross-spliced to vary in consonantal and vocalic properties.

1.2.2. Contextualizing Chinese Korean

The majority of ethnic Koreans currently living in China are descendants of immigrants who crossed the border between the mid 19th century and the end of the Second World War (Jin, 2008). The Korean communities in China have traditionally maintained strong ties to their Korean culture and language, including Korean-language education. However, what we refer to in this work as ‘Chinese Korean’ is far from a homogenous dialect. Inhabitants of different regions and cities are descendants of speakers of different Korean dialects. There is also influence from both standard North and South Korean: The North Korean standard (Pyongyang) was used as the model for Chinese Korean standardization in the mid-20th century (Tai, 2004), and since the establishment of diplomatic ties between China and South Korea in 1992, there has been an increase in exposure to South Korean culture and media, complemented by an increase in travel to South Korea by ethnic Koreans living in China. Furthermore, the influence of Mandarin has increased in recent years: The majority of speakers are bilingual (Jin, 2008), and there is a shift in dominant language use from Korean to Mandarin in some communities (e.g., Tai, 2004; Han, 2011; Han, 2014).

The current work focuses on two Chinese Korean communities in the border cities of Hunchun (Jilin province) and Dandong (Liaoning province) (**Figure 1**). Hunchun is smaller (~200,000) than Dandong (~800,000), but has a larger Korean population: It is located within the Yanbian Korean Autonomous Prefecture, and 36% of the residents



Figure 1: Korean communities in Hunchun and Dandong, China, are the target populations reported in this work. Map modified from an image from Google Maps.

are ethnic Koreans (China Data Center, 2006). On the other hand, Dandong has a smaller Korean community of around 20,000 (Cui, 2011). There is high exposure to Seoul Korean in both cities, both via South Korean media and visits to South Korea (details given below). Specific information about the language background of our participants will be reported in the Methods section; in the following section, we discuss some of the linguistic characteristics of the two dialects.

1.2.3. Dialectal differences in production of the stop contrast

As discussed above, there is considerable dialectal variability in production of the stop contrast, both in terms of relative use of specific cues and in whether or not there is an ongoing change in cue use. The varieties examined in the current study, dialects spoken in the cities of Hunchun and Dandong, stem from two distinct North Korean dialects (Hamkyeong dialect for Hunchun speakers and Phyeongan dialect for Dandong). One difference between these two dialects is that Hamkyeong (and in turn, Hunchun) maintains the lexical pitch accent distinction of Middle Korean that was lost in the Western dialects, including Phyeongan/Dandong, as well as Seoul (Ito & Kenstowicz, 2017). Below we discuss previous work reporting production patterns in Hunchun, Dandong, and cognate dialects (Hamkyeong and Phyeongan). We also show graphs of production patterns from the same Hunchun and Dandong participants of the current work, along with a matched group of Seoul participants, in **Figure 2**. This data is analyzed in Kang, Schertz, and Han

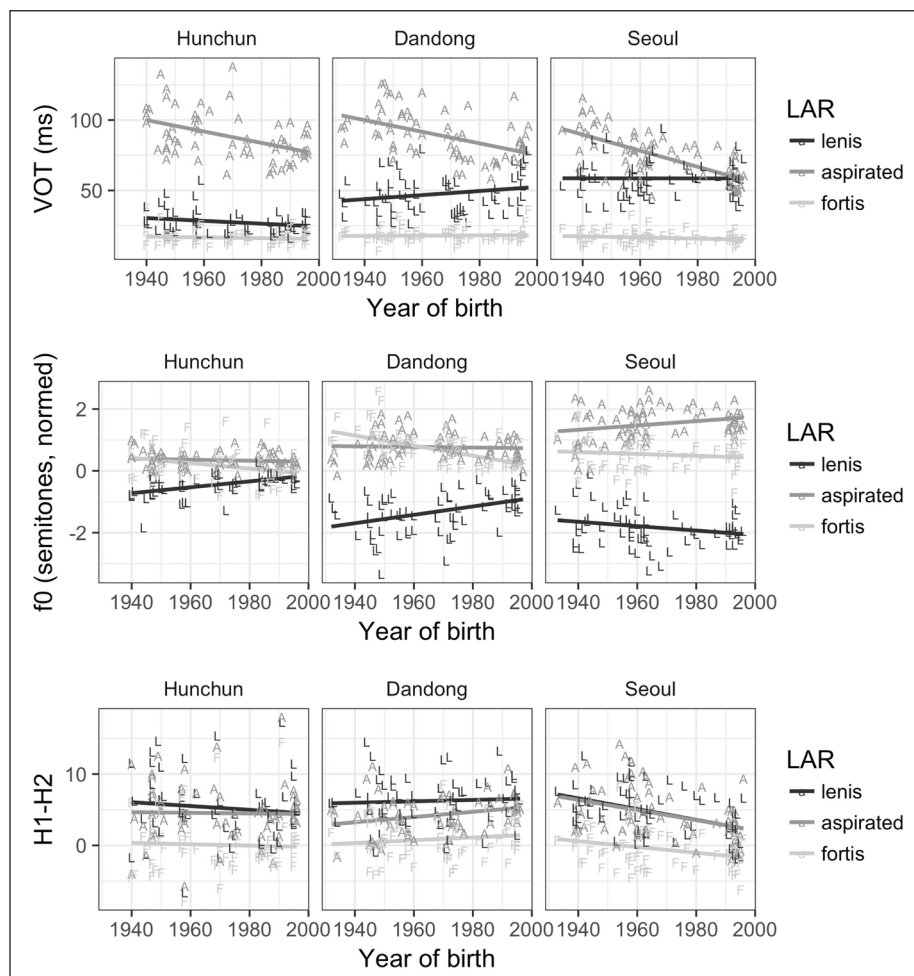


Figure 2: Data from 159 speakers from Hunchun, Dandong, and Seoul born in the 1930s-1990s (from Kang et al., forthcoming). Each data point represents a by-participant mean for a given laryngeal category, and the lines show the best-fit regression line. (a) VOT, (b) f0, (c) H1-H2.

(forthcoming), and full statistical results are available in the supplementary materials corresponding to the current article.

Hunchun/Hamkyeong: Previous studies on various varieties of Hamkyeong Korean have shown large and consistent differences from present-day Seoul patterns, with fortis and lenis stops both showing very short VOT values, in contrast to very long VOTs for aspirated stops (Ito & Kenstowicz, 2008; Chung, 2011; Kang & Han, 2012; Kang et al. forthcoming). In terms of f_0 , lenis stops are to be systematically lowest, and aspirated stops highest, but the difference is quite small, and can vary based on pitch-accent status (Kang et al., forthcoming). In terms of voice quality, all work on Hamkyeong Korean shows substantially lower H1-H2 for fortis than for lenis and aspirated stops, similar to Seoul Korean, and Kang and Han (2012) further showed that speakers from Qingdao, China (who speak a variety of Hamkyeong Korean) produced lenis stops with higher H1-H2 than aspirated stops. This small but systematic effect was mirrored in the data of Kang et al. (forthcoming) in Hunchun speakers, but a direct cross-dialectal statistical comparison showed no significant differences between use of H1-H2 in Hunchun versus Seoul.

Dandong/Phyeongan: Previous studies on Phyeongan Korean have been mixed. Jin (2008) and Jin and Silva (2017) found VOT patterns similar to those seen in Seoul, including an ongoing merger between lenis and aspirated stops, based on a production study of thirty-five 18–74-year-old speakers in Shenyang, the capital of Liaoning province. However, in Dandong (a different city in the same province), while there is evidence of an ongoing VOT merger, it is not as far advanced as the one in Seoul, such that Dandong speakers' VOT values for lenis and aspirated stops being better separated, and f_0 values being less well-separated, than their age-matched counterparts in Seoul (Kang & Han, 2012; Kang et al., forthcoming). In terms of voice quality, Kang and Han (2012) and Kang et al. (forthcoming) found, as expected, the lowest H1-H2 values for fortis stops, as well as systematically lower H1-H2 for aspirated than for lenis stops.

In order to formulate predictions for the current work, we summarize the use of cues in terms of each two-way contrast separately, and how this differs across dialects. These generalizations are based on previous production studies, as well as production data from the current participants, as reported in Kang et al. (forthcoming) and available in the supplementary materials.

Lenis-fortis contrast: The fortis and lenis categories are highly overlapping in both VOT and f_0 for Hunchun speakers, as in other Hamkyeong varieties (Kang et al., forthcoming). Dandong speakers have greater separation between the two categories on both dimensions, and the categories are even further separated in Seoul speakers. The relative overlap of the two categories in Hunchun suggests that Hunchun listeners may rely more heavily on other cues, such as voice quality of the following vowel, to make the distinction. A greater use of H1-H2 in Hamkyeong versus Seoul speakers was found by Kang and Han (2012), and this idea was also supported by patterns of a single speaker reported in Ito and Kenstowicz (2008); however, in production data from Hunchun, Dandong, and Seoul participants reported in Kang et al. (forthcoming), there were no apparent dialectal differences in use of H1-H2 for the lenis-fortis contrast.

Lenis-aspirated contrast: This contrast is the focus of most of the work done on dialectal and diachronic variability in the three-way contrast. VOT values for the two categories are well-separated in Hunchun speakers, almost totally overlapping in Seoul, and intermediately separated in Dandong. The use of f_0 goes in the opposite direction: There is substantial overlap in Hunchun, in contrast to complete separation in Seoul, with an intermediate level of separation in Dandong. In terms of H1-H2, there is evidence of

higher values corresponding to lenis versus aspirated stops in Hunchun and Dandong (Kang & Han, 2012; Kang et al., forthcoming), but this difference is very small.

In terms of within-dialect variability, no age-related differences have been found in Hunchun speakers except a small decrease in VOT for aspirated stops (Kang et al., forthcoming), but in both Dandong and Seoul, there is more overlap in VOT, and less in f_0 , in younger than in older speakers; nevertheless, the younger Dandong speakers do not show as much VOT overlap as older Seoul speakers (Kang & Han, 2012; Kang et al., forthcoming). Finally, in Seoul, females showed more merged VOT than their male counterparts (Oh, 2011; Kang, 2014; Kang et al., forthcoming). No consistent gender differences were found in Hunchun or Dandong.

Fortis-aspirated contrast: There is no clear dialectal variation in the fortis-aspirated contrast between Hunchun and Dandong listeners. In Seoul, fortis stops have, on average, a slightly lower f_0 than aspirated stops, but the two categories remain well-separated in terms of VOT, so we expect consistently primary use of this cue in all dialects. We also expect voice quality to play a role, as this separates the cues in all dialects.

1.3. The current study

This study presents a systematic investigation of perceptual cue weighting in two dialects of Korean which have been shown to differ both in use of cues in production and in the existence of a sound change in progress (where Dandong shows a clear sound change and Hunchun does not). We explore whether these dialectal and age-related differences are mirrored in perception, and whether listeners modify their perceptual strategies depending on the (apparent) dialectal affiliation of the talker.

We test listeners' use of VOT, f_0 , and voice quality in perception of the stop contrast via a series of identification experiments, where stimuli have been manipulated to vary orthogonally on these three dimensions. Using a novel analytical method in an effort to avoid problematic aspects associated with analysis of the three-way contrast in previous work, we first examine how listeners' phonetic category space along these dimensions differs by dialect and age. We then test whether listeners' knowledge of dialectal variation affects perception by comparing participants' responses across two conditions, one in which they are told that the talker is a member of their local dialect community, and the other in which the talker is said to be from Seoul.

Following previous work, we assume that the primary perceptual cues to the stop contrast are VOT and f_0 . However, our design also includes a manipulation of the baseline vowel (taken from natural productions of fortis and nonfortis vowels), in order to systematically assess the role that vocalic information, most notably voice quality, may play in listeners' perception of the contrast. Voice quality has been shown to serve as a cue in production (e.g., Cho et al., 2002) and perception (Kim et al., 2002; Kong, Beckman, & Edwards, 2011). However, in previous perception studies, voice quality was not manipulated separately from f_0 , making it difficult to assess its contribution independently of f_0 .

1.3.1. Predictions

Listener-level influences: Based on the idea that perceptual patterns correspond to dialect- and age-based community-level production norms, we expect less reliable distinction of the *fortis-lenis* contrast in Hunchun listeners in the VOT- f_0 space, and we expect that they may in turn rely more on other cues to the distinction, such as the voice quality of the preceding vowel, as proposed by Ito and Kenstowicz (2008). We also expect that Hunchun and Dandong listeners will differ in the relative position of *lenis and aspirated* categories in the perceptual space, with Hunchun listeners using primarily VOT to perceive the distinction, and Dandong relying more heavily on f_0 . Finally, we expect to see a reduced

reliance on VOT in younger Dandong listeners, corresponding to the fact that the lenis-aspirated contrast appears to be undergoing an ongoing VOT merger in the current participants' productions, and, under the view that the loss of a VOT distinction might lead to adaptive enhancement of f_0 (Kirby, 2013), we might also expect a concurrent increase in reliance on f_0 .

Talker-level influences: We also test the idea that listeners pay attention to sociophonetic variation and adjust their perception based on social characteristics of the talker. We predict that listeners will increase their reliance on f_0 as a cue to the lenis-aspirated contrast when listening to a talker they believe to be from Seoul, since Seoul speakers show even more extreme (high) use of f_0 and (low) use of VOT in production of the contrast. Finally, there is also dialectal variation in production of the lenis-fortis contrast. If Hunchun or Dandong listeners are aware of the fact that lenis stops are produced with a longer VOT in Seoul, they may show different perception patterns of the fortis-lenis contrast when listening to a talker from Seoul. Finally, following the prediction of Drager (2011), we might expect that older listeners, particularly those in Dandong, who have been present throughout the time course of sound change, might be more sensitive to the social patterning of acoustic cues and therefore show a larger influence of the apparent dialect of the talker.

2. Methods

2.1. Participants

We report data from 123 Hunchun and Dandong speakers, balanced for age and gender (**Table 1**). All speakers learned Korean as their first language. For the purposes of this study, we group participants into two age groups: Older (born before 1970) and Younger (born 1970 or later).² Data from eight additional participants were omitted because they did not complete all of the tasks reported in the current work ($n = 5$), had lived in multiple dialect communities for a substantial amount of time ($n = 1$), self-reported low proficiency in Korean ($n = 1$), or showed anomalous response patterns ($n = 1$).

Participants' answers to background questions about exposure to South Korean media and visits to South Korea are reported in **Table 1**. Participants self-reported how often they were exposed to South Korean media, on a scale of 1 (never), 2 (sometimes), 3 (commonly), 4 (often), and 5 (every day). Note that all groups report levels of exposure greater than 3, with relatively small within-group variance, pointing to a high level of exposure to South Korean media for participants in all age/dialect groups. We also asked

Table 1: Means (standard deviations in parentheses) of participants' age in 2015, self-reported amount of exposure to South Korean media, on a scale of 1 (never) to 5 (every day), and number of months spent in South Korea. Five speakers indicated that they had traveled to South Korea 'often': these are indicated separately in the final column of the table.

Dialect	Age Group	Number	Age in years	South Korean media exposure	Months spent in South Korea
Hunchun ($n = 59$)	Older	29 (17 F)	63 (8)	4.0 (0.3)	19.2 (5.5)
	Younger	30 (15 F)	29 (8)	4.6 (0.2)	1.0 (0.3) (+2 'often')
Dandong ($n = 64$)	Older	33 (20 F)	64 (9)	4.6 (0.2)	12.6 (4.8) (+3 'often')
	Younger	31 (14 F)	31 (10)	3.3 (0.3)	4.6 (2.4)

² The grouping was done via a median split. In exploratory data analyses, we also examined patterns using age as a continuous variable, and results were not substantively different from those found using Age as a categorical factor.

participants to report the amount of time spent in South Korea; as shown in **Table 1**, older listeners report having spent more time in South Korea than younger listeners in each dialect group; however, there is a large amount of within-group variability in terms of visit duration.



2.2. Overall design

Experiments were conducted in a quiet hotel room in Hunchun or Dandong, China, in the summer of 2015. This experiment was one of a larger set of perception and production tasks. Participants completed a three-alternative forced-choice perception task where they were presented with monosyllables (e.g., /tʌ/) varying on the dimensions of f₀ and VOT. On each trial, listeners chose which of three Korean stops (fortis /tʰ/, lenis /t/, or aspirated /tʰ/) they had heard.³ All participants completed two ‘Dialect Label’ conditions (Local and Seoul) that were identical in procedure but differed in the information given about the talker’s dialect (**Table 2**). In order to make the dialect label manipulation more plausible, listeners heard a male talker in one condition and a female in the other. In the Local Talker condition, listeners were told that they would be listening to a speaker of their own dialect (Hunchun or Dandong), while in the Seoul Talker condition, they were told that the talker was from Seoul. In both conditions, the Dialect Label manipulation was reinforced by visual and orthographic information (multiple cues to dialectal affiliation were used in order to maximize the salience of the perceived dialect manipulation). Half of the listeners heard the male talker in the Local condition and the female talker in the Seoul condition (Group A), while the other half heard the female talker in the Local condition and male talker in the Seoul condition.

2.3. Dialect Label manipulation

We provided listeners with several sources of information about the (intended) dialect of the talker, both explicit and implicit. First, in the instructions (which were both explained verbally by the experimenter and presented in writing at the beginning of each experimental block), listeners were told that they would be listening to a talker from their

Table 2: Overview of experiment design and cues to Dialect Label. In condition, listeners were given information cueing either a talker from their city (Local Dialect Label) or Seoul (Seoul Dialect Label). Each listener heard a male talker in one condition and a female in the other.

		Local Dialect Label Condition			Seoul Dialect Label Condition		
Listener city	Group	Talker	Dialect Label	Response options	Talker	Dialect Label	Response options
Hunchun (n = 59)	A (n = 30)	male	Hunchun	도, 또, 토 /to, tʰo, tʰo/ (Chinese Korean vowels)	female	Seoul	더, 떠, 터 /tʌ, tʰʌ, tʰʌ/ (Seoul vowels)
	B (n = 29)	female			male		
Dandong (n = 64)	A (n = 34)	male	Dandong		female		
	B (n = 30)	female			male		
Sample screenshots of the experiment presentation							

³ While these syllables can be real words, feedback from participants suggests that it is unlikely that the listeners considered the choices to be real words.

city (either Hunchun or Dandong) in the Local Dialect Label condition, or that they would be listening to a talker from Seoul in the Seoul Dialect Label condition. A photograph of the appropriate city was present on the screen throughout the experiment, providing a visual cue to the relevant dialect (cf. Hay & Drager, 2010).

Dialectal differences in realization of vowels in Chinese Korean versus Seoul dialects allowed for an additional, implicit cue to dialectal affiliation. Our nonword stimuli included a vowel whose acoustic properties map to different phonological categories in the different dialects: /o/ (Korean ‘ㅏ’) in Hunchun and Dandong but /ʌ/ (Korean ‘ㅓ’) in Seoul Korean. In other words, listeners would have heard this vowel used in words like 어른, /ʌlin/, ‘adult,’ in their local dialect, but the same vowel used in words like 오래, /oɛ/, ‘long time,’ in Seoul Korean. In our design, the auditory stimuli were identical across the two Dialect Label conditions, but the orthography shown in the response choices was different, corresponding to the dialect-specific vowels. In the Local Dialect Label condition, listeners were asked to choose between 도, 또, and 토, which correspond to /to, tʰo, tʰo/, while in the Seoul Dialect Label condition, listeners chose between 더, 떠, 터, corresponding to /tʌ, tʰʌ, tʰʌ/. This dialect-specific orthographic/phonological mapping of the stimuli provided additional, implicit information about dialect mode.⁴

2.4. Stimuli

2.4.1. Baseline token creation

Auditory stimuli for the task consisted of monosyllabic CV syllables (e.g., /tʌ/), manipulated from natural productions. Baseline tokens for stimulus manipulation were created from recordings of two talkers’ productions of /tʌ, tʰʌ, tʰʌ/. The female talker was from Seoul and born in 1982; the male was from Asan (Chungnam province) and was born in 1940.

We chose two natural tokens for each talker, one fortis and one lenis. Fortis and lenis (as opposed to aspirated) were chosen as baselines because they have been shown to have the greatest difference in voice quality: Cho et al. (2002) found that fortis has the lowest, and lenis has the highest H1-H2. We spliced aspiration from an aspirated consonant onto the vocalic portions of these productions to create four baseline tokens. We manipulated the formants of these baseline tokens to match those formant values judged to be a good exemplar of /o/ in Chinese Korean and /ʌ/ in Seoul discussed above. Formant manipulations were done following the instructions for source-filter synthesis in Praat version 5.3.82 (Boersma & Weenink, 2014); formant values were set to remain stable throughout the entire vowel.

Given that we have only two talkers who differ in age, gender, and potentially other characteristics, it is not possible to attribute talker-related differences found in the results to any one of these specific characteristics. Therefore, we refrain from interpreting effects of the ‘Talker’ factor; however, we do include and report all effects of this factor in order to highlight potential talker-related differences in perception results.

2.4.2. Acoustic manipulations

The four baseline tokens were then manipulated on the dimensions of f0 and aspiration duration, using the PSOLA algorithm (Moulines & Charpentier, 1990) in Praat.

⁴ In order to ensure that our stimuli contained a vowel that did in fact fall into different orthographic/phonological categories in the different dialects, we asked native speakers of each dialect to perform a goodness rating task on a series of stimuli varying in vowel quality (F1 and F2). For our final stimuli, we used the formant values from a vowel that was considered a good exemplar of /o/ by the Chinese Korean raters and a good exemplar of /ʌ/ by the Seoul raters. The formant values from these tokens (male: F1 = 533 Hz, F2 = 833 Hz; female: F1 = 700 Hz, F2 = 1133 Hz) were used to create the baseline stimuli.

While the purpose of this was to serve as an additional cue to dialectal affiliation, it should be noted that it also creates an additional difference between the two Dialect Label conditions. While we do not see any theoretical reason why this should affect results, this is a confound in the design.

Fundamental frequency (f0): We varied f0 at vowel onset in a five-step series. We chose the endpoints by calculating each talker's f0 range at vowel onset in their production of all coronal stops, then extending the range by 0.25, resulting in endpoints of 200 Hz (min) and 335 Hz (max) for the female talker and 135 Hz (min) and 200 Hz (max) for the male talker. For each value of f0, f0 was set at vowel onset, remained steady for two-thirds of the vowel duration, then fell linearly to 175 Hz (female talker) or 120 Hz (male talker).

Aspiration: We varied aspiration duration in an eight-step series, from 0 to 100 ms. Aspiration duration was manipulated by increasing or decreasing the period of aspiration in the baseline token using the PSOLA algorithm in Praat. A 7 ms burst always preceded the aspiration duration, so the VOT of our stimuli ranged from 7 to 107 ms.

The final set of stimuli consisted of 80 tokens for each talker (5 steps of f0 \times 8 steps of aspiration \times 2 baseline vowels [fortis and nonfortis]). Listeners heard two randomized repetitions of stimuli from a single talker in each condition, for a total of 320 trials (in both conditions combined) per listener.

2.5. Procedure

Participants were given both oral and written instructions before each task and completed several practice trials before beginning the main experiment. Participants were told that they would hear a talker from their city (Hunchun or Dandong) in the Local Dialect Label block, or from Seoul in the Seoul Dialect Label block. They were asked to choose which of the three Korean stops best represented the sound they heard (**Table 2**). The choices were visible on the screen throughout the whole experiment. All participants completed the Local block before the Seoul block. The tasks were administered on a Microsoft Surface 3 tablet. Participants heard the stimuli through headphones (Sony MDR7506) and responded by tapping their choice on the touch screen. The experiments were administered by research assistants who were proficient L2 Korean speakers (L1 Mandarin).

2.6. Analysis

Our primary research goal is to explore variability in listeners' perception of the Korean three-way stop contrast. We start from the premise, based on a large body of previous work, that f0 and VOT are primary cues used to distinguish the contrast, and our analysis examines how listeners' categories differ in the VOT-f0 perceptual space. **Figure 3** shows three 'maps' of different hypothetical perceptual spaces, schematizing the range of variability found in our perception data in terms of the relative positions of the lenis-aspirated categories, where we expect to see the most variation. (a) represents a listener who distinguishes the contrast primarily based on VOT (i.e., an idealized older Hunchun listener), and (c) represents a listener who distinguishes aspirated from lenis

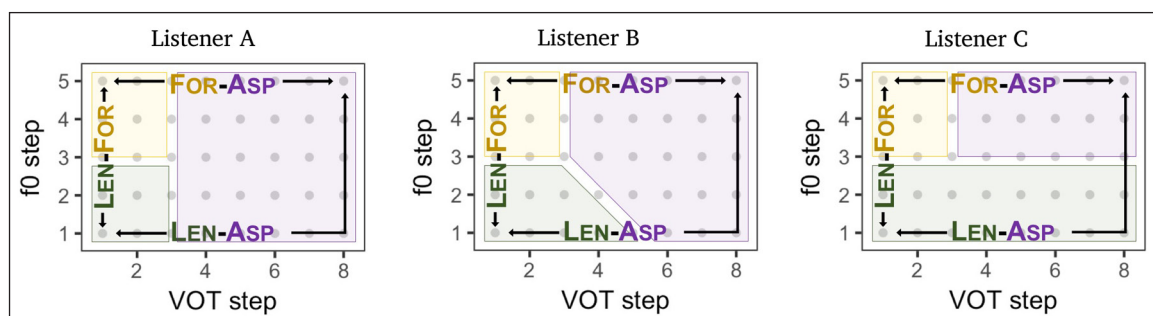


Figure 3: Hypothetical categorization of the three-way Korean stop contrasts on the dimensions of VOT and f0 by three listeners. The labels on the extremes of the space indicate the contrast primarily differentiated by that 'dimension.' The three examples below schematize the variability found in our perception data (details in-text).

primarily based on f_0 (i.e., an idealized younger Dandong listener), while (b) represents an intermediate pattern.

Logistic regression provides an attractive way to analyze the relative role of various factors in predicting listeners' responses in a forced-choice task. Most previous work on perception of the three-way stop contrast has used logistic regression with VOT and f_0 as predictor variables, interpreting a larger effect for a given variable as more 'use' of that acoustic cue by listeners (e.g., Kong et al., 2011; Lee et al., 2013; Schertz, Cho, Lotto, & Warner, 2015). However, trying to account for three-way forced choice data with a binary analysis poses problems. Analyzing perception of each category separately (e.g., fortis versus other, aspirated versus other, lenis versus other) as in Lee et al., 2013, results in redundancy, since the three analyses are not independent in the context of a forced-choice experiment (a change in one category necessarily implies a change in the adjacent category). On the other hand, dividing the data into three two-way contrasts (as in individual analyses in Schertz et al., 2015) results in artefacts of the 'missing' category in the analyses. A pair of two-way analyses (Helmert-coded as in Kong et al., 2011 or the multinomial regression analysis in Schertz et al., 2015) resolves the issues above but does not allow for parallel comparison of all three two-way contrasts.

Instead of using f_0 and VOT as independent predictors of responses across the entire acoustic space, we choose subsets of the stimulus space where we expect listeners to show a binary contrast, and focus only on the relevant region for analysis of each binary contrast. Specifically, we analyze each two-way contrast along a single acoustic continuum: the subset of the VOT- f_0 space extrema where we expect to find the category boundary for the relevant contrast. These subsets are shown in **Figure 4**. For example, we expect the high- f_0 extreme (the top of the plots) to contain only fortis and aspirated responses (see **Figure 3**). Furthermore, we expect the choice of fortis versus aspirated along this continuum to be predictable as a function of VOT, with categorically fortis responses at the lower endpoint of the continuum, and categorically aspirated responses at the higher endpoint. Therefore, in our analysis of the fortis-aspirated contrast, we model listeners' choice as a function of continuum step (VOT in this case) as well as our other predictor variables of interest. Similarly, we expect the low-VOT extreme of the

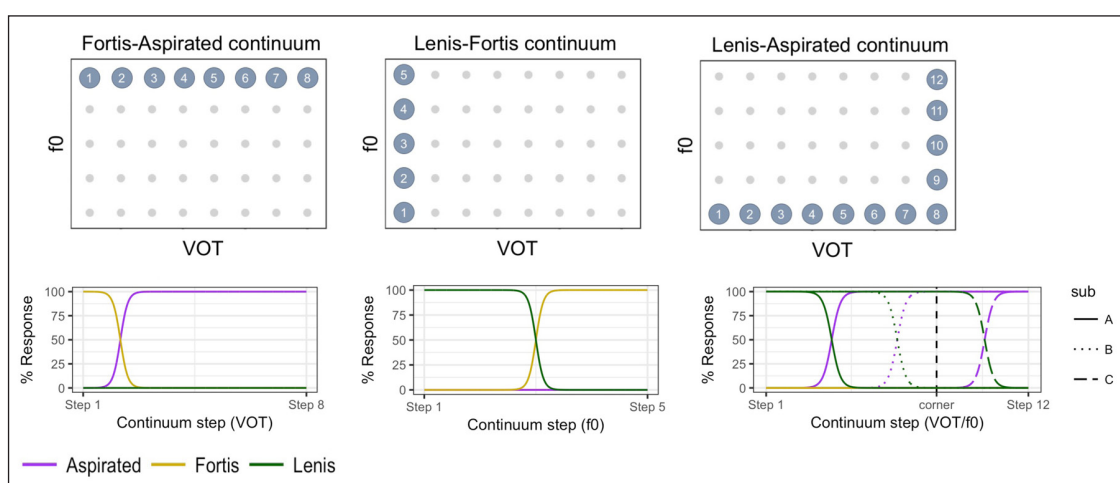


Figure 4: The top row shows the three continua of stimuli from the extremes of the VOT- f_0 space that were used to analyze perception of each of the three two-way contrasts. The figures below show hypothetical categorization curves along each of these continua. The three listeners schematized above in Figure 3 differ in their categorization of the lenis-aspirated 'hybrid' continuum; their different patterns are indicated by different line types. The vertical dashed line in the lenis-aspirated continuum indicates the stimulus (step 8) at the bottom right corner.

acoustic space (left side of the plot) to subsume the lenis-fortis contrast, ranging across low to high values of f_0 .

The lenis-aspirated contrast differs from the two other pairwise contrasts in that it is not possible to choose a single one-dimensional VOT or f_0 continuum on the extremes of the acoustic space that will encompass the complete categorization curves for all listeners, because of the aforementioned dialectal differences in the relationship between these two categories. For some listeners (**Figure 3a**), the low- f_0 region appears to be the relevant series (ranging from lenis at the low-VOT endpoint to aspirated at the high-VOT endpoint), while for others (**Figure 3c**), the high-VOT region is the relevant series (ranging from lenis at the low- f_0 endpoint to aspirated at the high- f_0 endpoint). For the listeners who show patterns intermediate between (a) and (c), we would expect to see a boundary further on the VOT continuum (**Figure 3b**), or lower on the f_0 continuum (not shown). Given this, to test the change in listeners' lenis-aspirated boundary, we adopt a 'hybrid continuum,' which ranges across the VOT range in the low- f_0 region, then continues with increasing values of f_0 in the high-VOT region (i.e., starting at the left side of the bottom axis, turning the corner, and continuing up the right axis).

2.6.1. Validating the hybrid continuum

As the hybrid continuum is somewhat unconventional, we wanted to ensure its appropriateness for its use in our analysis, as well as its interpretability. First, we examined listeners' lenis and aspirated categorization curves across the two one-dimensional portions of the continuum (VOT ranging across the low- f_0 region, and f_0 ranging across the high-VOT region) separately. If the hybrid dimension is appropriate, listeners should show a category boundary on *one* of the two dimensions, but not both. We verified that this was indeed the case by estimating individual listeners' 50% crossover points using individual logistic regression models across each of the two portions of the continuum separately. The complementary distribution of individual listeners' crossover points across the two individual continua therefore validates the use of a single hybrid continuum that subsumes the full distribution of listener category boundaries.

We also wanted to ensure that findings from this analysis can be interpreted in a meaningful way. Traditionally, the differences between the hypothetical listeners (a) and (c) in **Figure 3** is described as a difference in the relative use of VOT and f_0 . Therefore, the category boundary on our hybrid continuum should correlate with a more standard measure of relative VOT versus f_0 use across the whole stimulus space. In order to test this, we calculated two values for each participant: 1) the 50% crossover point on our hybrid continuum, and 2) the relative use of VOT and f_0 across the whole acoustic space.⁵ These two values were highly correlated ($r = .93$), indicating that our hybrid continuum is a reasonable proxy for relative use of f_0 and VOT as estimated by a more conventional method. In addition to validating the interpretability of this hybrid dimension, this strong correlation indicates that although our analysis only uses a subset of the acoustic space, it is a good predictor of performance across the entire space.

To sum up, our analysis quantifies variability in Korean stop perception by examining listeners' response data across three acoustic continua, each subsuming a separate two-way contrast. This allows us to analyze the three contrasts in a parallel manner, without redundancy, which is not possible using more standard logistic regression analyses.

⁵ 50% crossover points were calculated via a logistic regression model predicting aspirated response from continuum step, using the subset of data included in the hybrid continuum. Relative VOT/ f_0 weights were calculated via a logistic regression model, using data from the whole stimulus space, predicting aspirated response from VOT and f_0 ($\text{asp.choice} \sim \text{VOT} + f_0$), then transforming the VOT and f_0 coefficients such that they summed to zero, as in Schertz et al., 2015. This can be conceptualized as the slope of the line in the VOT* f_0 space that best separates the lenis and aspirated categories.

2.6.2. Statistical analysis

As discussed above, we analyze listeners' phonetic categorization patterns along three continua that have been selected to subsume the three pairwise contrasts. For each continuum, we expect responses to follow a prototypical pattern for a binary forced-choice categorization task, similar to the idealized curves in **Figure 4**. For example, for the fortis-aspirated continuum, we expect to see 100% fortis/0% aspirated responses at the low endpoint, and 0% fortis/100% aspirated responses at the high endpoint, with no lenis responses at all in this region.

We analyze each continuum separately using mixed-effects logistic regression, using the *lmer* package in R (Bates, Maechler, Bolker, & Walker, 2015), with listeners' choice of the high-endpoint category as the response variable: For example, in analysis of the fortis-aspirated continuum, the binary response variable is 'aspirated' response (versus 'other,' either lenis or fortis). Under the idealized circumstances laid out above, changes in this response variable correlate inversely with the 50% crossover point of a given category (e.g., more aspirated responses correspond to an earlier 50% crossover point in the aspirated curves in **Figure 4**). If our assumptions about listeners' behavior are correct, we therefore expect the intercept of our regression models to correlate with the crossover point, allowing us to then turn to how this intercept/crossover point differs as a function of our factors of interest. Because we find crossover point to be more intuitive than units of log odds in the interpretation of results, we present most effect sizes and visualizations in terms of crossover points (in units of step difference) rather than units of log odds.

Model structure: Our primary analysis consisted of three mixed-effects logistic regression models, one for each continuum. The response variable was the high-endpoint category for each continuum: aspirated for the lenis-aspirated and fortis-aspirated continua, and fortis for the lenis-fortis continuum. Each model contained the same predictor variables, all simple-coded (-0.5, 0.5) (reference level in italics): Listener Dialect (*Hunchun* vs. *Dandong*), listener Age (*Older* vs. *Younger*), baseline Vowel (*nonfortis* vs. *fortis*), Talker (*female* vs. *male*), and perceived Dialect Label (*Local* vs. *Seoul*). Continuum Step (VOT step for the fortis-aspirated continuum, f0 step for the lenis-fortis continuum, and VOT/f0 step for the hybrid lenis-aspirated continuum) was included as a covariate and was centered prior to analysis: We did not include interactions of continuum with the other factors because we expected talker- or listener-dependent variations to occur mainly in the boundary location rather than in the slope of the categorization curve.

We followed model selection procedures in Matuschek, Kliegl, Vasishth, Baayen, and Bates (2017) to select optimal random effects structures for each model, testing models with random by-subjects intercepts as well as random slopes for all within-subjects variables against models with incrementally fewer slopes using the *anova* function in R, with an alpha-level of 0.2. We then used the same selection criteria for the inclusion of interactions of fixed effects, starting with full (five-way) interactions and testing them against models excluding the highest-order interactions. In the case of significant interactions, we performed follow-up tests using the *phia* package (De Rosario-Martinez, 2015) to test whether the effect of interest held at each level of the other factors.

3. Results

3.1. Perception of the three-way contrast in the VOT-f0 space

Figure 5 shows overall perception patterns across the acoustic space used in our experiments: Stimuli varied in VOT, f0, and baseline vowel (fortis and nonfortis). These heat plots allow for visualization of the overall patterns in the same dimensionality as the production data. However, it is difficult to see subtle differences, so we show here only

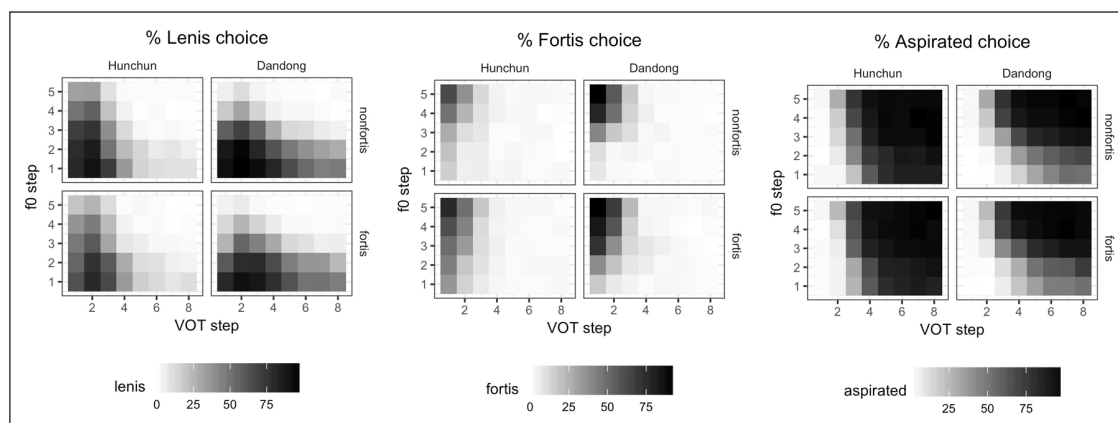


Figure 5: Overall perception results, with darkness showing the percentage of listeners' responses for (a) lenis, (b) fortis, and (c) aspirated stops in the VOT- f_0 space. In each plot, separate subplots are shown for Hunchun (left panel) and Dandong (right panel) listeners, and for nonfortis (top panel) and fortis (bottom panel) baseline vowels separately.

the factors that had the largest influence on our results: Listener Dialect and baseline Vowel. Responses are broken down by all other factors in detail using categorization curves in the subsequent sections.

We focus first on the two primary dialect-related differences seen in the production data in **Figure 2**. First, we saw that Hunchun speakers showed more overlap in the VOT- f_0 space than Dandong speakers in production of the fortis-lenis contrast. We can see a similar pattern in these perception graphs: The low-VOT region (left side of the plots) shows more of a mixture of lenis and fortis responses for Hunchun listeners, as compared to Dandong, for whom the category boundary is somewhat more clear-cut. Comparing the perception of stimuli across the fortis and nonfortis vowel conditions, we see that the baseline vowel does influence categorization in perception: Focusing on the middle graph (% Fortis choice), we see that there are overall slightly more fortis than nonfortis responses when the vowel is fortis (bottom panel) in both dialects. However, this difference is small and clearly not a categorical cue in either dialect: If it were, we would find *no* fortis responses in the nonfortis vowel condition.

The second main dialectal difference in the production data was the relative placement of the lenis-aspirated contrast in the VOT- f_0 space, with Hunchun speakers producing primarily a VOT contrast (with little to no f_0 contrast), while Dandong speakers' productions contrasted more in f_0 , and less in VOT. We see this pattern mirrored in the perception results most clearly by looking at the Aspirated plot: In Hunchun, Aspirated responses are most clearly defined by long VOT, while in Dandong, stimuli with long VOT are not necessarily categorized as aspirated, and this can be attributed to the fact that the lenis category extends across the low- f_0 region, even when VOT is long, in Dandong speakers.

Overall, then, it appears that the main trends in production are reflected, at least to some degree, in perception. We now turn to a detailed examination of each pairwise contrast, and how perception differs based on all of our factors of interest.

3.1.1. Perception of the fortis-aspirated contrast

Results of listeners' perception of the fortis-aspirated continuum (i.e., those stimuli with the highest values of f_0) are shown in **Figure 6** and **Table 3**. The graphs show listeners' responses across the fortis-aspirated continuum (spanning the range of VOT values along the highest step of f_0), with line types showing how responses compare across each level

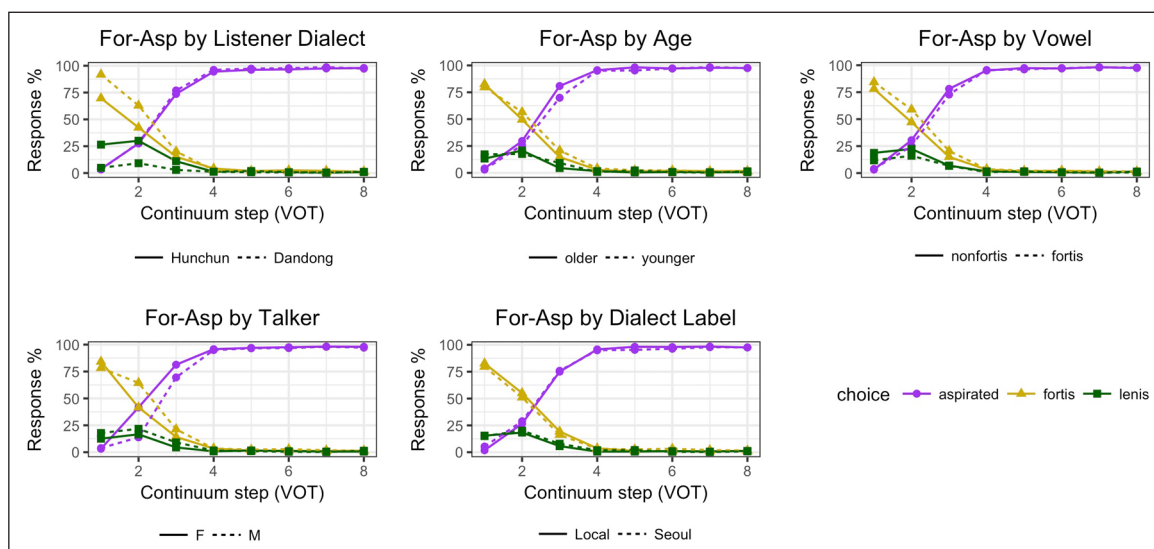


Figure 6: Categorization curves along the fortis-aspirated continuum, broken down by each predictor variable: (a) Listener Dialect, (b) Age, (c) Vowel, (d) Talker, (e) Dialect Label.

Table 3: Output of statistical model predicting aspirated response along the fortis-aspirated continuum. The reference level for each factor is in italics.

	Estimate	SE	z value	p value
Intercept	4.676	0.292	15.999	<.001***
Continuum Step	2.287	0.124	18.490	<.001***
Listener Dialect (<i>Hunchun</i> versus <i>Dandong</i>)	0.061	0.176	0.348	0.728
Age (<i>older</i> versus <i>younger</i>)	-0.342	0.176	-1.946	0.052.
Vowel (<i>nonfortis</i> versus <i>fortis</i>)	-0.197	0.088	-2.240	0.025*
Talker (<i>female</i> versus <i>male</i>)	-1.019	0.143	-7.148	<.001***
Dialect Label (<i>Local</i> versus <i>Seoul</i>)	-0.018	0.125	-0.144	0.886

*** $p < .001$; ** $p < .01$; * $p < .05$; . $p < 0.1$.

of our factors of interest. Our only prediction regarding this contrast was that we expected fortis vowels to elicit more fortis responses; we did not expect any differences in perception of this contrast based on listener or talker characteristics.

We first observe general patterns which hold across all factors, then examine each factor separately. As expected, aspirated responses range from 0% at the lower endpoint to 100% at the high endpoint of the continuum, and listeners primarily chose fortis when not choosing aspirated. However, note that at the low end of the continuum there was a small but substantial percentage of lenis responses, which we had not expected. In other words, our assumption that listeners’ decisions along this continuum are constrained to a binary choice does not always hold at the lower end of the continuum. Nevertheless, our response variable as it is formulated (‘Aspirated’ versus ‘Other’) does behave as expected across all conditions.

Our logistic regression model predicts aspirated response (the purple lines in **Figure 6**) as a function of Continuum Step (in this case, VOT step), Listener Dialect, Age, Vowel, Talker, and Dialect Label. Based on the model selection procedures laid out above, the optimal model included random slopes for Continuum Step and Talker, and the model reported below with no interactions in the fixed effects did not differ significantly from a model with all interactions. As expected, there was a significant effect of continuum step, indicating an increase in Aspirated response across the continuum, and the significant intercept indicates an overall higher probability of Aspirated response.

Listener-level effects: There was no significant main effect for Listener Dialect, shown in the corresponding graph by the similar shape of the Aspirated response curves across the two dialects. Overall, listeners of the different dialects and age groups do not differ in their perception of the Fortis-Aspirated contrast.

Vowel effects: There was a main effect of Vowel, with fortis vowels having a negative coefficient, corresponding to fewer aspirated responses and a slightly higher category boundary; in other words, the fortis category extended further on the continuum when the baseline vowel was fortis. However, this effect is very small, with an average crossover difference of only 0.10 steps.

Talker effects: There was a main effect of Talker, with a lower coefficient for the male talker representing a 0.51 step higher boundary for the male than the female talker.

Dialect Label effects: The effect of Dialect Label was not significant.

3.1.2. Perception of the lenis-fortis contrast

Results of listeners' perception of the lenis-fortis continuum (i.e., those stimuli on the left edge of the VOT-f₀ space) are shown in **Figure 7** and **Table 4**. The graphs show listeners' responses across the lenis-fortis continuum, with line types showing how responses compare across each level of our factors of interest. As expected, we see increasing fortis and decreasing lenis responses across the continuum, and there are practically no aspirated responses. However, the curves are flatter than the idealized categorization curves we would expect to see, and they do not reach floor or ceiling at the endpoints. Furthermore, the slopes of the curves differ by condition: For example, Dandong listeners' responses appear to be more categorical (i.e., showing a steeper slope) than those of Hunchun listeners. We discuss implications of this, and do a follow-up analysis, after our primary analysis (see Section 3.2).

Our logistic regression model predicts fortis response (the yellow lines in **Figure 7**) as a function of Continuum Step, Listener Dialect, Age, Vowel, Talker Voice, and Dialect Label. Based on the model selection procedures laid out above, the optimal model included random slopes for Continuum Step, Vowel, and Talker Voice, and three-way interactions in fixed effects. As expected, there was a significant effect of continuum step, indicating an increase in fortis response across the continuum.

Listener-level effects: There were no significant main effects of Listener Dialect or Age, which we take to indicate a lack of any major effects primarily determined by

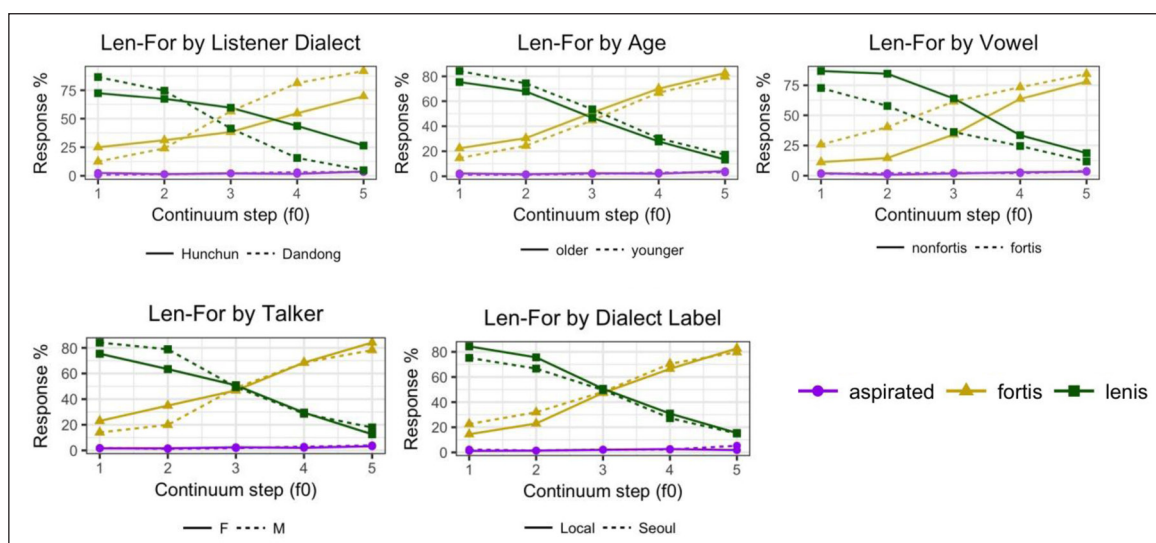


Figure 7: Categorization curves along the lenis-fortis continuum, broken down by each predictor variable: (a) Listener Dialect, (b) Age, (c) Vowel, (d) Talker, (e) Dialect Label.

Table 4: Output of statistical model predicting fortis response along the lenis-fortis continuum. The reference level for each factor is in italics.

	Estimate	SE	z value	p value
Intercept	-0.074	0.130	-0.568	0.570
Continuum Step	1.244	0.070	17.885	<.001***
Listener Dialect (<i>Hunchun vs. Dandong</i>)	0.258	0.364	0.709	0.478
Age (<i>older vs. younger</i>)	-0.452	0.249	-1.817	0.069.
Vowel (<i>nonfortis vs. fortis</i>)	1.367	0.112	12.182	<.001***
Talker (<i>female vs. male</i>)	-0.510	0.136	-3.748	<.001***
Dialect Label (<i>Local vs. Seoul</i>)	0.296	0.132	2.243	0.025
Listener Dialect * Age	0.593	0.491	1.207	0.227
Listener Dialect * Vowel	-0.549	0.253	-2.172	0.030*
Listener Dialect * Talker	-0.607	0.374	-1.622	0.105
Listener Dialect * Dialect Label	-0.680	0.265	-2.560	0.010*
Age * Vowel	0.098	0.186	0.528	0.597
Age * Talker	-0.058	0.267	-0.216	0.829
Age * Dialect Label	0.152	0.260	0.586	0.558
Vowel * Talker	-1.203	0.165	-7.280	<.001***
Vowel * Dialect Label	-0.356	0.164	-2.172	0.030*
Talker * Dialect Label	-0.484	0.491	-0.986	0.324
Listener Dialect * Age * Vowel	0.162	0.370	0.438	0.662
Listener Dialect * Age * Talker	1.471	0.528	2.787	0.005**
Listener Dialect * Age * Dialect Label	-0.607	0.516	-1.177	0.239
Listener Dialect * Vowel * Talker	-0.086	0.330	-0.260	0.795
Listener Dialect * Vowel * Dialect Label	-0.259	0.329	-0.786	0.432
Listener Dialect * Talker * Dialect Label	-0.569	1.012	-0.563	0.574
Age * Vowel * Talker	0.115	0.329	0.351	0.726
Age * Vowel * Dialect Label	0.840	0.329	2.552	0.011*
Age * Talker * Dialect Label	2.786	0.984	2.832	0.005**
Vowel * Talker * Dialect Label	0.358	0.372	0.964	0.335

these listener-level factors. However, both Listener Dialect and Age showed significant interactions with other factors, as discussed below.

Vowel effects: There was a significant main effect of Vowel, with a fortis vowel eliciting more fortis responses, corresponding to a 1.37 step decrease in category boundary when compared to the nonfortis vowel condition. There were significant two-way interactions of Vowel with Listener Dialect, Talker, and Dialect Label, as well as a significant three-way interaction between Vowel, Age, and Dialect Label. The magnitude of the effect of Vowel broken down by these interacting factors is shown in **Figure 8**. The graph shows the difference in predicted 'fortis' response (in terms of log odds) between baseline fortis and nonfortis vowels for each subgroup. The direction of the effect was the same across all subgroups (with fortis vowels eliciting more fortis responses), but groups differed in the extent of this effect. Hunchun listeners made (slightly) more use of the baseline vowel than Dandong listeners, and Vowel influenced listeners' responses for the female more than for the male talker. In terms of the three-way interaction, there was a significant

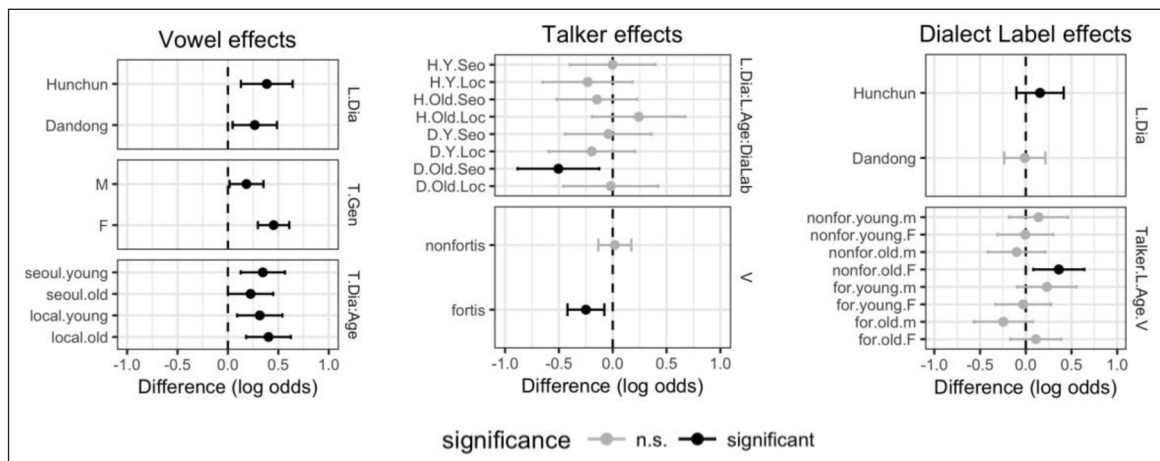


Figure 8: Interaction effects for the lenis-fortis contrast. In each plot, the point shows the predicted difference in log odds for ‘fortis’ choice for each level of the relevant factor (Vowel in (a), Talker in (b), and Dialect Label in (c)), broken down by subgroups that interacted significantly with the factor. For example, both Dandong and Hunchun listeners were more likely to respond fortis for fortis than nonfortis baseline vowels, as represented by the positive effect size, but this vowel-based difference is of greater magnitude in Hunchun than in Dandong. Error bars show one standard error based on model predictions (*testInteractions* function in the *phia* package in R, De Rosario-Martinez, 2015). Black lines indicate a significant effect of the relevant factor (e.g., Vowel in the first panel) for each subgroup.

difference in the use of Vowel by Dialect Label for older, but not for younger, listeners, with older listeners using Vowel more in the Seoul than in the Local condition.

There were main effects of both Talker and Dialect Label, as well as significant interactions of these two factors with others, as shown in **Figure 8**. However, as can be seen in the interaction plots, there was not a clear or consistent effect of either of these factors.

Talker effects: The main effect for Talker shows that the female talker elicited more fortis responses than the male talker, and follow-up tests suggest that this effect is driven by fortis vowels: When the vowel was fortis, there were more fortis responses for females than males. Put together with the findings from Vowel above, this corresponds to an overall greater apparent use of voice quality for the female talker⁶. However, in terms of the interaction with Listener Dialect, Age, and Dialect Label, the effects do not appear consistent: Only one subgroup showed significant effects (the female talker elicited more fortis responses for older Dandong listeners in the Seoul condition). Given the inconsistency of the effects, we do not interpret this further.

Dialect Label effects: We did not have any predictions regarding Dialect Label but found a main effect of Dialect Label, which indicates that overall, the Seoul condition elicits more Fortis vowels than the Local condition. The effect appears to be driven by Hunchun listeners, while there is no effect for Dandong listeners. Broken down by Vowel, Age, and Talker, no consistent Dialect Label effects emerge. We therefore do not interpret these further.

3.1.3. Perception of the lenis-aspirated contrast

Results of listeners’ perception of the lenis-aspirated continuum are shown in **Figure 9** and **Table 5**. The graphs show listeners’ responses across the hybrid lenis-aspirated continuum (ranging across low- f_0 and continuing up the high-VOT dimension), with line types showing

⁶ While it could be the case that listeners systematically use voice quality more for female than male talkers (or younger rather than older talkers), it is likely simply the case that in our specific stimuli, there were more cues to voice quality in the female than the male talker. We cannot distinguish between these possibilities with the current design.

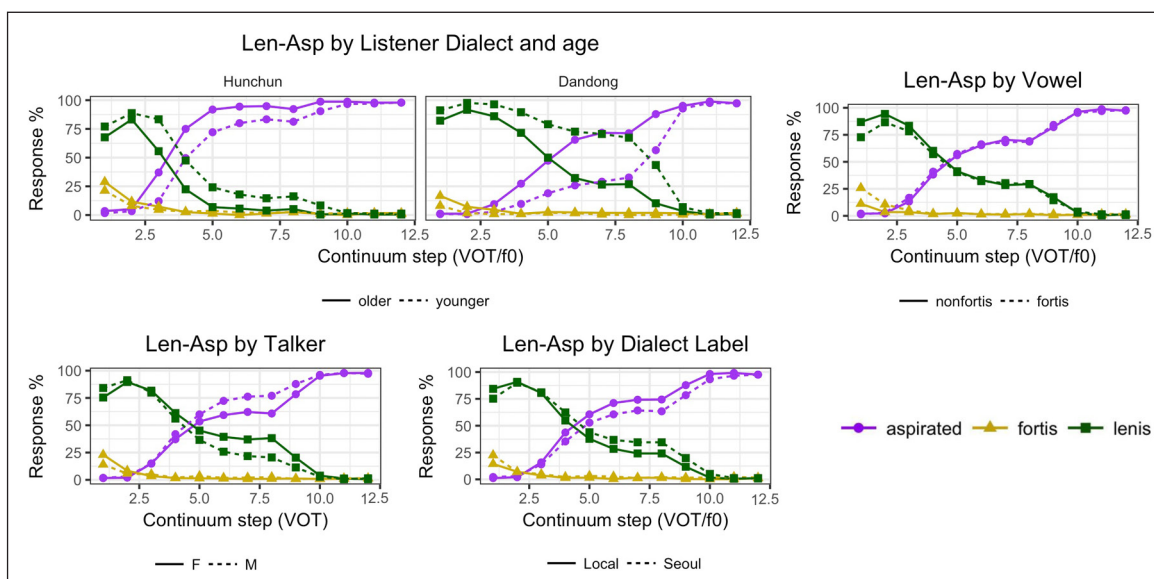


Figure 9: Categorization curves along the hybrid lenis-aspirated continuum, broken down by each predictor variable (for this contrast, the effect of age is shown separately for Hunchun and Dandong listeners because the effect of age is substantially different in the two dialects).

Table 5: Output of statistical model predicting Aspirated response along the lenis-aspirated continuum. The reference level for each factor is in italics.

	Estimate	SE	z value	p value
Intercept	1.448	0.195	7.434	<.001***
Continuum Step	1.219	0.043	28.529	<.001***
Listener Dialect (<i>Hunchun</i> versus <i>Dandong</i>)	-2.866	0.300	-9.549	<.001***
Age (<i>older</i> versus <i>younger</i>)	-1.863	0.299	-6.223	<.001***
Vowel (<i>nonfortis</i> versus <i>fortis</i>)	0.089	0.072	1.233	0.218
Talker (<i>female</i> versus <i>male</i>)	0.708	0.165	4.289	<.001***
Dialect Label (<i>Local</i> versus <i>Seoul</i>)	-0.728	0.168	-4.344	<.001***
Listener Dialect * Age	-1.437	0.519	-2.771	0.006**
Listener Dialect * Vowel	0.048	0.142	0.335	0.738
Listener Dialect * Talker	0.478	0.307	1.558	0.119
Listener Dialect * Dialect Label	0.075	0.319	0.233	0.815
Age * Vowel	0.146	0.142	1.029	0.304
Age * Talker	0.296	0.306	0.966	0.334
Age * Dialect Label	-0.669	0.319	-2.097	0.036*
Vowel * Talker	-0.244	0.135	-1.798	0.072.
Vowel * Dialect Label	0.089	0.135	0.659	0.510
Talker * Dialect Label	0.513	0.578	0.889	0.374

how responses compare across each level of our factors of interest. As expected, we see an increase in Aspirated responses across the continuum, and a corresponding decrease in Lenis responses. At the beginning of the continuum, there are some (unexpected) Fortis responses, more so for Hunchun listeners. This reflects what we have seen earlier: Some listeners have categorically fortis responses across the whole low VOT range. As this has

already been accounted for in the lenis-fortis analysis, we focus on aspirated responses, where we see the full range of responses (0% to 100%).

Our logistic regression model predicts aspirated response (the purple lines in **Figure 9**) as a function of Continuum Step (on the hybrid VOT/f₀ continuum), Listener Dialect, Age, Vowel, Talker, and Dialect Label. Based on the model selection procedures laid out above, the optimal model included random slopes for all within-subjects factors (Continuum Step, Vowel, Talker, and Dialect Label), and two-way interactions in fixed effects. As expected, there was a significant effect of continuum step, indicating an increase in aspirated response across the continuum. The positive intercept indicates more aspirated responses than chance, as indicated by the crossover point below the midpoint of the continuum.

Listener-level effects: Listener Dialect, Age, and their interaction were significant, and Age also interacted with Dialect Label. Dandong listeners had a higher crossover point (i.e., fewer aspirated responses, corresponding to more use of f₀ relative to VOT) than Hunchun listeners, and within each group, younger listeners had a higher crossover point than older listeners. The significant interaction indicates that the age effect is larger in Dandong than in Hunchun (though follow-up tests show it is significant in both dialects). In terms of the interaction with Dialect Label, the age effect is significant in both Dialect Label conditions, though slightly greater in the Seoul condition; this interaction will be accounted for in the Talker-level effects section below.

Vowel effects: The main effect of Vowel was not significant, nor were its interactions with other factors, indicating that the voice quality of the vowel does not influence the lenis-aspirated category boundary.

Talker effects: There was a main effect of Talker, with the female talker eliciting fewer aspirated responses (i.e., a higher crossover point, corresponding to more use of f₀) than the male talker. Talker was not involved in any significant interactions.

Dialect label: There was also a main effect of Dialect Label, with the Seoul talker eliciting fewer Aspirated responses (i.e., higher crossover point/more use of f₀) than the Local talker. The aforementioned interaction of Dialect Label with Age indicates that the effect is larger for younger than older listeners. Follow-up tests showed that the Dialect Label effect was significant ($p < .001$) for younger listeners, and not significant ($p = .086$), but numerically in the same direction for older listeners.

3.1.4. Interim summary and interpretation

In order to compare the magnitude of overall effects, **Figure 10** shows the absolute values of the difference in crossover points between the two levels of each factor. The color of the bars indicates the significance of the factor: Black bars indicate overall significance of the main effect, either without interactions, or, if there are interactions with other factors, significance holds at each level of the other factor. Grey bars indicate that the factor interacted with one or more other factors, and that while the direction of effect was consistent, it was not always significant when broken down by other factors. White bars indicate that the factor showed inconsistent or no significant effects.

Fortis-Aspirated contrast: The only factor predicted to affect the fortis-aspirated category boundary along the VOT dimension was Vowel, and it did, in the expected direction. However, the effect with the greatest magnitude for the fortis-aspirated contrast was one that was not predicted, Talker. The fortis category extended further along the VOT dimension for the males. Perception of the fortis-aspirated contrast did not appear to differ based on listener-level characteristics or Dialect Label.

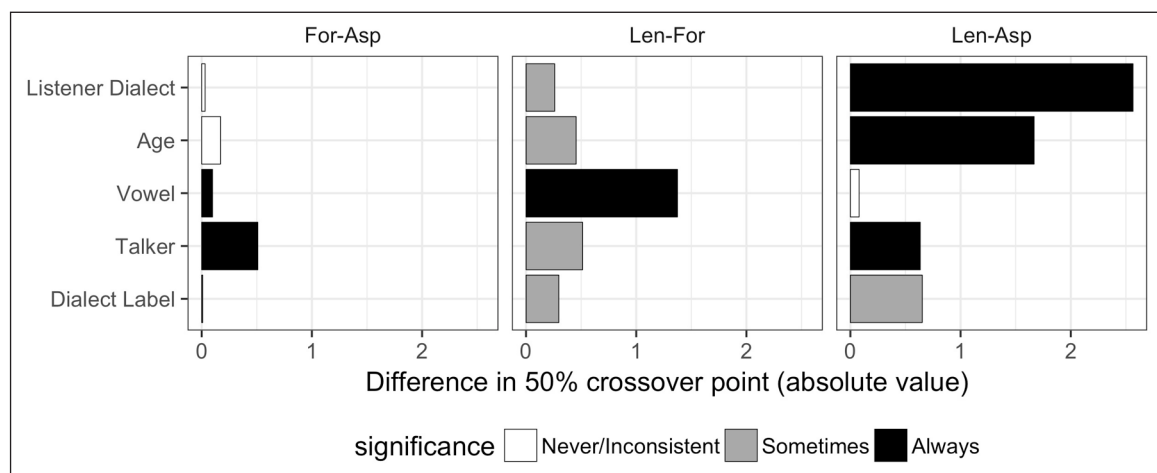


Figure 10: Differences in 50% crossover points between the two levels of each predictor variable. Units are continuum steps. Bar colors represent significance (see details in-text).

Lenis-Fortis contrast: As expected, listeners used information in the following vowel to inform their categorization across the lenis-fortis continuum, and Hunchun listeners used this information more than Dandong listeners. There was also more use of vowel for the female than for the male talker, which appeared to be driven by the fact that the female's fortis vowels elicited more fortis responses, as opposed to a symmetrically larger contrast. Hunchun, but not Dandong, listeners showed a small Dialect Label difference, with the Seoul condition eliciting more Fortis responses than the Local condition. There were not consistent overall differences based on Listener Dialect or Age.

Lenis-Aspirated contrast: As predicted, there were strong and consistent dialect and age-related differences in perception of the lenis-aspirated contrast, with Dandong listeners showing higher category boundaries (corresponding to heavier reliance on f_0 than VOT) than Hunchun listeners. The same pattern holding for younger versus older listeners, and an interaction indicated a greater effect of age for Dandong listeners. There was a consistent influence of Talker on responses, with more use of f_0 for the younger, female talker. Finally, Dialect Label showed an overall effect in which the Seoul label elicited a higher category boundary than the Local label, but this was only significant for younger listeners.

3.2. Follow-up: Dialectal differences in predictiveness of f_0 versus voice quality in the lenis-fortis contrast

Our choice of continua assumes that the primary dimensions influencing the category boundary are VOT (for fortis-aspirated), f_0 (for lenis-fortis), and combined VOT- f_0 (for lenis-aspirated). That this assumption holds can be seen clearly from most of the graphs: We see floor and ceiling categorization effects at the ends of the relevant continuum, with only a small difference in category boundary for other factors. However, as discussed above, this is not the case for the lenis-fortis contrast. We found a strong effect of vowel, and the relatively shallow slope of the curves implies that although f_0 was an important predictor of Fortis versus Lenis response, this expected primary cue was not as 'primary' as the dimensions covered by the continua for the other two contrasts.⁷

⁷ We did not test the effect of VOT in predicting fortis-lenis categorization. However, based on the overall perception graphs (Figure 5), it seems clear that VOT is not predictive for this contrast in Hunchun, and while it is predictive in Dandong, we think this is more attributable to the lesser use of VOT for the lenis-aspirated contrast in Dandong.

We therefore performed a direct test of the predictiveness of f_0 as compared to baseline vowel in categorization of the lenis-fortis contrast. We built three models predicting Fortis response from 1) f_0 Continuum Step, 2) Vowel, and 3) both f_0 Step and Vowel. Listener Dialect was also included as a fixed factor in both models since we found differences in use of the dimensions across the two dialects, and random by-subjects slopes and intercepts were included. It is not possible to directly compare the effect size of the two predictors given that they are on different scales; therefore, we quantified predictiveness by comparing how well each model predicted listeners' responses. **Table 6** gives confusion matrices showing how model predictions compared with listeners' responses, as well as d' , a measure of accuracy.

We focus on the d' scores as a metric of the predictability of the models. As expected from the previous results, the vowel-only model was more successful in predicting responses of Hunchun listeners than Dandong listeners (70% versus 63%), while the f_0 continuum step model was more successful for Dandong than Hunchun listeners (82% versus 74%). The fact that the accuracy of the vowel model was lower for both dialects suggests that f_0 is more predictive than vowel in the lenis-fortis distinction, even for Hunchun listeners. The combined models for each dialect were more accurate than either of the individual models, as expected; however, the model was still substantially more successful for Dandong than Hunchun listeners (84% versus 77%). Therefore, it appears that Hunchun listeners' responses are generally not well-explained by the parameters manipulated in this work, nor by age or talker information examined here, as evidenced by the fact that they did not show consistent significant effects in the main analysis above.

4. Discussion

4.1. Summary of main findings

We approached this study with hypotheses stemming from two assumptions: 1) the listener's perception is guided by dialectal and age-based production norms, and 2) listeners adjust their perceptions based on the apparent dialect of the talker. Overall, the results of the experiments show that both listener- and talker-related characteristics do

Table 6: Confusion matrices comparing model predictions (rows) with listener responses (columns). Results are shown from models with predictor variables of Listener Dialect and a) Vowel, b) f_0 Continuum step, and c) both Vowel and f_0 step. For each listener response type (fortis or lenis), the two rows of cells show the percentage of time the model classified that trial as fortis or lenis. Percentage 'correct' response (where the model prediction matched listener response category), as well as d' scores are shown.

	Hunchun											
	Vowel model				f_0 model				Vowel + f_0 model			
	For	Len	% Correct	d'	For	Len	% Correct	d'	For	Len	% Correct	d'
Pred. For	655	325	70.2	1.04	690	262	74.4	1.28	752	260	77.1	1.45
Pred. Len	378	1002			343	1065			281	1067		
	Dandong											
	Vowel model				f_0 model				Vowel + f_0 model			
	For	Len	% Correct	d'	For	Len	% Correct	d'	For	Len	% Correct	d'
Pred. For	971	549	63.3	0.65	1140	228	82.4	1.87	1151	197	84.1	2.03
Pred. Len	391	649			222	970			211	1001		

affect perception, in the direction predicted by our hypotheses; however, in both cases, we also found unpredicted effects or asymmetries that raise further questions.

For the first, ‘listener-level,’ premise, we found that the two major dialectal differences seen in Hunchun and Dandong production patterns, reported in Kang et al. (forthcoming), had analogs in perception. First, f_0 was indeed less predictive of Hunchun listeners’ than Dandong listeners’ responses, reflecting the fact that the lenis-fortis contrast showed more f_0 overlap in Hunchun than in Dandong productions. Second, dialectal differences in production of the lenis-aspirated contrast (greater f_0 and smaller VOT differences in Dandong versus Hunchun talkers), were also reflected in perception, as evidenced by a higher category boundary for Dandong listeners on a ‘hybrid continuum’ that was designed to quantify the relative use of the two dimensions. Finally, an age-based difference in perception of the lenis-aspirated contrast was predicted for Dandong listeners, where production data suggests a change in progress. This prediction was supported in that there was an age effect in Dandong, with younger listeners showing increased reliance on f_0 relative to VOT. However, an age-related difference in the same direction, albeit of a smaller magnitude, was found in Hunchun as well, which was not predicted given that there is no apparent sound change in Hunchun production. This unexpected finding is discussed further below.

For the Dialect Label manipulation, we expected that listeners’ responses would be affected by their expectations, based on their experience with dialectal differences, namely that Seoul speakers make more use of f_0 in production of the lenis-aspirated contrast than Dandong or Hunchun speakers. In support of this, listeners’ category boundaries indicated greater use of f_0 when listening to the Seoul talker, although this was only significant for younger listeners. We also considered the possibility of the influence of Dialect Label on the lenis-fortis contrast, given dialectal differences in production of this contrast. We found that Hunchun listeners showed more fortis responses, i.e., a lower category boundary along the f_0 continuum, in the Seoul than the Local talker condition. We give one possible explanation for this result below.

Finally, we found two Talker-related differences. First, in the fortis-aspirated contrast, the older, male talker elicited a higher category boundary (i.e., more fortis responses) along the VOT continuum. Second, in the lenis-fortis contrast, the younger, female talker elicited more fortis responses than the male, but only when the vowel was a baseline fortis vowel.

4.2. Methodological contributions and limitations

This study presented a novel method for analyzing the three-way contrast, using logistic regression to predict responses on three continua chosen to represent the end-points of each of the three two-way contrasts. Overall, this method met our goal of avoiding redundancy and analytical artefacts present in previous analyses, and we hope that it will be used and further evaluated in future work. One apparent limitation is the fact that it does not use all of the data that was collected (i.e., the full acoustic space). We showed that it nevertheless captured the same patterns that would be shown using a more standard analysis technique.

One complication in the interpretation of this study is the fact that each listener heard a different talker in the two Dialect Label conditions (i.e., either the female talker in the Local condition and the male talker in the Seoul condition, or vice versa). This choice was made intentionally: In order to avoid this, we would either have to have the same talker across both conditions, which we thought would detract from the plausibility of the Dialect Label manipulation, or expose listeners to only one of the Dialect Label conditions, which would make it so that we could not look at within-listener patterns across the

conditions. The fact that we alternated which talker was heard in each condition allows us to test for an independent effect of Dialect Label. Nevertheless, Dialect Label is not a strictly within-subjects factor, which would be preferable from a design perspective, and which would allow for effects to be seen more clearly, without the additional variability due to Talker. While this was a necessary limitation given the scope and population of the current study, future work looking directly at the effect of apparent dialect should consider ways to avoid this design confound.

The use of two talkers resulted, not unexpectedly, in differences in perception patterns. In principle, differences in the perception of two given talkers could be due to 1) uncontrolled acoustic differences between the voices or 2) listeners' use of social expectations. This confound is particularly clear in cases like the current one, where there are two talkers with obviously different characteristics (gender, age): The voices show considerable acoustic differences, and they also vary along salient social dimensions. In principle, any apparent effect of Talker could be due to listeners' use of different strategies when listening to male versus female talkers (or younger versus older talkers). However, in this particular case, we think it is likely that these two results are due to acoustic differences between the two talkers, as opposed to use of social expectations by the listeners. For the fortis-lenis difference, acoustic analysis of the two talkers' stimuli supports our conjecture that the female's fortis token may have simply sounded 'more fortis': we found a larger difference in H1-H2 between the fortis and nonfortis baseline tokens for the female than the male (difference in H1-H2 is 4.9dB for the female versus 2.9dB for the male talker), and the female's fortis token has a much lower H1-H2 than the male (-3.4dB vs. 0.6dB). For the fortis-aspirated VOT difference, we do not have a specific explanation for why this should vary by talker, but it is known that the VOT boundary can vary based on talker- or stimulus-specific characteristics (e.g., Lisker & Abramson, 1967), and given that the fortis and aspirated categories are very well-separated in terms of VOT, there is ample room for variation.

Although we think this explanation is plausible, the variability introduced by the talkers is a limitation in interpretation of our results. However, we believe that this Talker-driven variation in responses also highlights an important gap in the phonetic cue-weighting literature. Most studies that manipulate stimuli along one or more acoustic dimensions use only a single talker. While this evades the issue of talker variability, using a single talker also seriously limits the generalizability of the results. Given what we know about the influence of both acoustic and social information on perception, we believe that it is important for work going forward to consider talker variability explicitly, even if it is not a primary question of interest. Finally, despite the fact that the use of two different talkers introduces an additional source of variability, the effect of Dialect Label, the target of one of our primary hypotheses, is still able to be interpreted as an independent effect, since it varied orthogonally with respect to the two talkers.

4.3. Use of voice quality in perception of the Korean laryngeal contrast

While most previous work looking at dialectal differences in perception of the laryngeal contrast has focused on VOT and f_0 , in this work we showed that listeners from different dialects also vary in their use of voice quality, or other spectral information present in the vowel. The baseline vowel manipulation used in this study provides the first systematic investigation into the independent role of voice quality in perception of the Korean stop contrast, supporting findings by Kim et al. (2002) (Chang, 2013 showed a similar secondary use of voice quality in perception of the Korean fricative laryngeal contrast). Our results showed that, as expected, voice quality of the following vowel played influenced listeners' perception, with fortis vowels eliciting both a small increase in fortis responses

along the fortis-aspirated continuum (0.2 steps), and a more substantial increase along the lenis-fortis continuum, with a fortis vowel lowering the threshold for a fortis response by almost a whole continuum step (0.9 steps). Voice quality did not affect perception of the lenis-aspirated contrast, which was expected given the similar breathier (high H1-H2) quality of vowels following lenis and aspirated stops, which may also underlie inconsistent findings in previous production studies (e.g., Cho et al., 2002; Ahn, 1999). The two base vowels in this study represented the two ends of the voice quality continuum and were therefore not expected to capture the smaller voice quality difference between lenis and aspirated stops. A systematic examination of the acoustic and perceptual correlates of voice quality as it is used in the laryngeal contrast, going beyond the most widely-used metric of H1-H2, is a topic for future work.

One of our expectations was that Hunchun listeners would make less use of f_0 , and rely more on voice quality, as predicted by a claim about Yanbian Korean, a related dialect, by Ito and Kenstowicz (2008), and by their production patterns, which show a high degree of overlap on the f_0 dimension. We found that Hunchun listeners do in fact make less use of f_0 and more use of voice quality than Dandong listeners. However, this dialectal difference is small, even smaller than the difference in voice quality use based on the talker. Furthermore, if Hunchun listeners showed a *primary* reliance on H1-H2, as predicted by Ito and Kenstowicz (2008), then a model with Vowel as a predictor should better predict listeners' responses than a model with f_0 as a predictor. However, our follow-up analysis found that Hunchun listeners still rely more on f_0 than on Vowel, and that the overall predictability of the two dimensions together is less for Hunchun than Dandong. In other words, while we found the expected effect of less use of f_0 by Hunchun listeners, we did not find a compensatory cue that they rely on instead. Whether listeners use other cues that were not considered in this study, whether the lack of results were a function of the specific task used in this study, or whether the contrast is simply less robust, therefore remains an open question.

4.4. Perceptual cue-weighting in the context of sound change

Production patterns show evidence of change in Dandong speakers' use of acoustic cues to the lenis versus aspirated stop contrast, in the form of an ongoing VOT merger, similar to the well-known sound change in Seoul. On the other hand, production of the contrast appears stable across age groups in Hunchun. Interestingly, while evidence of a sound change in production was only clearly present in Dandong, we found age-related differences in *perception* of the lenis-aspirated contrast in Hunchun as well, with younger listeners in both dialects showing decreased reliance on VOT relative to f_0 , in comparison with their older counterparts. Here we discuss several potential reasons for this finding.

One possibility is that there is an ongoing sound change in Hunchun. There are two age-related differences reported in Kang et al. (forthcoming) that could be interpreted as a change in progress: First, younger speakers' aspirated stops are shorter in VOT. This is unlikely to be driving a sound change, since lenis and aspirated stops are not approaching a merger. Another finding is that speakers are using less f_0 in pitch accent, which Lee and Jongman (2018) argue allows for greater use of f_0 in the laryngeal contrast. However, our Hunchun listeners if anything use *less* f_0 to signal the laryngeal contrast. Alternatively, there could be a sound change in Hunchun that is not captured by the type of production data (reading words in isolation) reported here. Bang, Sonderegger, Kang, Clayards, and Yoon (2015) found that both VOT merger and f_0 enhancement in the Seoul lenis-aspirated contrast disproportionately affect high-frequency words (which tend to be hypo-articulated or reduced). Similarly, Harrington, Kleber, Reubold, and Siddins (2015) found that the German tense-lax vowel distinction is more merged when de-accented (also a context for

hypoarticulation), and thus more likely to induce misperception. Examination of different speech styles more likely to show hypoarticulation might therefore reveal an incipient change in Hunchun production. In either of these cases, given that we *do* see change in other dialects with this methodology, the change is clearly at a much earlier stage in Hunchun.

Another possibility is that there is no sound change in production yet, but that there will be, under the idea that sound change in perception precedes that of production (Ohala, 1993; Hyman, 1976); i.e., a change in production comes about as a result of listeners misparsing intrinsic phonetic effects of f_0 as extrinsic, eventually phonologizing them. This account predicts that a sound change in Hunchun should occur in the future.

However, it is also possible that there is, in fact, no current, incipient, or impending change in Hunchun production norms. Rather, the age-related difference in Hunchun perception could be driven by exposure to change in other communities (including Seoul). Even in the absence of a community-level sound change in production, perception norms could shift because of changes in other communities. Younger listeners might show a greater effect because they are disproportionately exposed to speech from younger members of the other community. However, even assuming that the exposure is qualitatively equivalent for younger and older listeners, an asymmetry might be expected because the older listeners' representations are more influenced by their long-term experience with the 'older' version of the other community's speech norms. If it is indeed the case that there are differences in perception due to exposure to dialectal variation, then a more general implication is that people who are exposed to many different accents might have different perceptual patterns than those who have not, without necessarily showing any difference in production patterns.

The current results do not allow us to distinguish between these possibilities: 1) there is an incipient or forthcoming sound change in production, or 2) production is stable, but perception is shifting. Either of these would have interesting implications. The first case would provide strong evidence that perception is leading, and perhaps predictive of, this sort of sound change. This could be tested by examining the Hunchun population across a longer period of time (or examining younger listeners), but the directionality of perception versus production in sound change could also be tested more generally via comparison of perception and production data from other dialects of Korean which are known to be undergoing sound change. In the second case, if the 'change' in perception is being driven by Seoul, it is worth noting that it also applies even in the Local condition, in which case exposure to dialectal variability is also affecting perception of listeners' 'own' dialect.

Our current results, therefore, bring up more questions than answers about the time-course of sound change in perception. However, we hope that the asymmetry between perception and production shown by Hunchun listeners in this work points to the need for the provenance behind this mismatch to be tested empirically.

4.5. Use of social expectations in speech perception

The effect of the apparent dialect of the talker on listeners' responses, with the Seoul talker eliciting more use of f_0 in the lenis-aspirated contrast than the talker said to be from the local community, is consistent with previous findings demonstrating an influence of top-down information on perception, reinforcing previous findings of talker effects on speech perception (e.g., Niedzielski, 1999; Strand & Johnson, 1996; Hay et al., 2006b).

Most work looking at the effect of perceived social information has focused on vowels, presumably because vowels are a primary source of dialectal variation in English. There has also been some work on fricatives (Dufour et al., 2014; cf. Chang, 2017). Moreover,

previous work has primarily focused on domains where there is meta-linguistic awareness of the variable in question. Many phonological and phonetic factors are familiar to Korean speakers as dialectal markers, including vowel quality, sibilant place of articulation, intonation, and speech rate (e.g., Lee, 2009; Jeon, 2011; Yang, 2013); however, the phonetic realization of stop contrasts has not been mentioned as a perceptually salient dialectal difference. Therefore, the use of talker information found in this work shows that listeners can make use of their exposure to dialectal variation even when the variation is below the level of conscious awareness.

The Dialect Label effect was only significant for younger listeners (although it was numerically in the same direction for older listeners). This age-based discrepancy cannot be attributed to less exposure to the dialectal variation by older speakers, since they have as much (self-reported) exposure to Seoul Korean as younger listeners. However, it is possible that the exposure is qualitatively different: Older speakers may be exposed to different types of speakers in the present, or their input may be influenced by longer experience with 'older' versions of Seoul Korean. Another possibility is that younger listeners are more sensitive to dialectal variability. In either case, this finding stands in contrast to the one other study of which we are aware that examines the interaction of listener age and use of social information: Drager (2011) found that older, not younger, listeners were more sensitive to a community-level change in progress. Together, these findings suggest that age-based differences in use of social information in speech perception may be an interesting avenue for future work to explore.

Although any talker-level differences need to be interpreted with caution, we discuss one here: an overall greater use of f_0 for the female than the male talker. We chose age-gender pairings that we hoped would represent extremes of a sound change in progress: a younger female (a group that might be expected to lead change and represent the most innovative production patterns) and an older male (who might be expected to represent a more conservative speech style). Listeners' behavior was in accordance with our prediction of more use of f_0 for speakers who would be expected to be at more advanced stages of the sound change. Therefore, we also interpret this as suggestive of listeners' use of their expectations about the speech of talkers based on their perceived age and/or gender. However, given the confounds inherent in using two different talkers, there are alternative possible explanations for this finding. Differences could also result from other unpredicted social expectations, and acoustic properties of the two voices could also lead to perceptual differences without having to reference social information at all. A study with balanced gender/age and multiple talkers from each age group is necessary to tease apart these factors more definitively.

4.6. Conclusion

This work explored the joint role of listener-level factors (listeners' own demographic information) and talker-level factors (listeners' expectations given social characteristics of the talker) on speech perception. We documented dialectal and age-based variation in how Korean listeners from two dialects used VOT, f_0 , and vocalic information to classify their native three-way stop contrast, the realization of which varies substantially across dialects and ages. While most group-level perception differences corresponded with dialectal differences seen in production, there was an interesting age-related asymmetry, an apparent change in perception in one of the dialects (Hunchun) where there is no evidence of a corresponding change in production. We also showed that listeners' perception differs across talkers, consistent with the hypothesis that younger listeners used top-down information about dialectal variation to inform their perception of stops. This highlights the importance of considering listener expectations, as well as both the social

and acoustic characteristics of the talker, when interpreting differences in performance on experimental tasks. We hope this will spur more specific investigations of how listener-level norms, talker-level expectations, and their interaction with listeners' experience shape perception.

Additional File

The additional file for this article can be found as follows:

- **Appendix.** Production data. DOI: <https://doi.org/10.5334/labphon.67.s1>

Acknowledgements

The authors would like to thank Professor Sun Ying at Liaoning University, Yunyan Luo, and Yuanyang Song for invaluable help with the data collection process. Kyeong-Hye Kim, Dong-Ki Han, Hae-Dong Park, Sung-Geol Kim, and Na-Young Ryu helped with stimulus preparation, and Rachel Soo and N.-Y. Ryu assisted with data processing. This research was supported by SSHRC Grant #435-2013-2092 to Yoonjung Kang.

Competing Interests

The authors have no competing interests to declare.

References

- Ahn, H. (1999). Post-release phonatory processes in English and Korean: Acoustic correlates and implications for Korean phonology. Doctoral dissertation, University of Texas, Austin.
- Bang, H., Sonderegger, M., Kang, Y., Clayards, M., & Yoon, T. (2015). The effect of word frequency on the timecourse of tonogenesis in Seoul Korean. *Proceedings of the International Congress of Phonetic Sciences*.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Package “lme4” v. 1.1.7. <https://cran.r-project.org/web/packages/car/index.html>.
- Boersma, P., & Weenink, D. (2014). Praat: Doing phonetics by computer, version 5.3.82: <http://www.praat.org>.
- Chang, C. (2013). The production and perception of coronal fricatives in Seoul Korean: The case for a fourth laryngeal category. *Korean Linguistics*, 15(1), 7–49. DOI: <https://doi.org/10.1075/kl.15.1.02cha>
- Chang, Y. (2017). The influence of dialect information on the perception of the Mandarin alveolar-retroflex contrast. *Concentric: Studies in Linguistics*, 43(1), 1–23.
- China Data Center. (2006). “China Data Online.” Ann Arbor: University of Michigan. <http://chinadataonline.org>. Accessed March 12, 2016.
- Cho, T., Jun, S., & Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *Journal of Phonetics*, 30, 193–228. DOI: <https://doi.org/10.1006/jpho.2001.0153>
- Chung, E. (2011). *An experimental study of vowels and consonants of Standard Yukjin dialect [Hyentay Yukcin Pangen Camoumey Kwanhan Silhemumsenghakcek Yen-gu]*. MA Dissertation, Seoul National University.
- Cui, J. (2011). Investigations and studies on the use of language of Korean nationality discourse community in Dandong area. Master's thesis, Bohai University.
- De Rosario-Martinez, H. (2015). Package ‘phia’ v. 0.1. <https://github.com/heliosdrm/phia>.
- D’Onofrio, A. (2015). Persona-based information shapes linguistic perception: Valley Girls and California vowels. *Journal of Sociolinguistics*, 19, 241–256. DOI: <https://doi.org/10.1111/josl.12115>

- Drager, K. (2010). Sociophonetic variation in speech perception. *Language and Linguistics Compass*, 4(7), 473–480. DOI: <https://doi.org/10.1111/j.1749-818X.2010.00210.x>
- Drager, K. (2011). Speaker age and vowel perception. *Language and Speech*, 54, 99–121. DOI: <https://doi.org/10.1177/0023830910388017>
- Dufour, S., Kriegel, S., Alleesaib, M., & Nguyen, N. (2014). The perception of the French /s/-/ʃ/ contrast in early Creole-French bilinguals. *Frontiers in Psychology*, 5. DOI: <https://doi.org/10.3389/fpsyg.2014.01200>
- Evans, B., & Iverson, P. (2004). Vowel normalization for accent: An investigation of best exemplar locations in Northern and Southern British English sentences. *The Journal of the Acoustical Society of America*, 115(1), 352–361. DOI: <https://doi.org/10.1121/1.1635413>
- Fridland, V., & Kendall, T. (2012). Exploring the relationship between production and perception in the mid front vowels of U.S. English. *Lingua*, 122, 779–793. DOI: <https://doi.org/10.1016/j.lingua.2011.12.007>
- Gordon, M., & Ladefoged, P. (2001). Phonation types: A cross-linguistic overview. *Journal of Phonetics*, 29(4), 383–406. DOI: <https://doi.org/10.1006/jpho.2001.0147>
- Han, S. (2011). The language identity of Korean-Chinese society in Qingdao, China. *Dialectology*, 14, 114–136.
- Han, S. (2014). The language change of Korean-Chinese society in China. *Korean Studies*, 32, 411–438. Center for Korean Studies, Inha University.
- Harrington, J., Kleber, F., Reubold, U., & Siddins, J. (2015). The relationship between prosodic weakening and sound change: Evidence from the German tense/lax vowel contrast. *Laboratory Phonology*, 6, 87–117. DOI: <https://doi.org/10.1515/lp-2015-0002>
- Hay, J., & Drager, K. (2010). Stuffed toys and speech perception. *Linguistics*, 48, 865–892. DOI: <https://doi.org/10.1515/ling.2010.027>
- Hay, J., Nolan, A., & Drager, K. (2006a). From fush to feesh: Exemplar priming in speech perception. *The Linguistic Review*, 23, 351–379. DOI: <https://doi.org/10.1515/TLR.2006.014>
- Hay, J., Warren, P., & Drager, K. (2006b). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, 34, 458–484. DOI: <https://doi.org/10.1016/j.wocn.2005.10.001>
- Hillenbrand, J., Cleveland, R. A., & Erickson, R. L. (1994). Acoustic correlates of breathy vocal quality. *Journal of Speech, Language, and Hearing Research*, 37(4), 769–778. DOI: <https://doi.org/10.1044/jshr.3704.769>
- Holliday, Jeffrey J., & Kong, E. (2011). Dialectal variation in the acoustic correlates of Korean stops. *Proceedings of the International Congress of Phonetic Sciences*, 17, 878–881.
- Hyman, L. Phonologization. (1976). *Linguistic studies offered to Joseph Greenberg*, 2, 407–418.
- Ito, C., & Kenstowicz, M. (2018). Pitch Accent in Korean. In *Oxford Research Encyclopedia of Linguistics*. DOI: <https://doi.org/10.1093/acrefore/9780199384655.013.242>
- Ito, Chiyuki, & Michael Kenstowicz. (2008). Mandarin loanwords in Yanbian Korean I: Laryngeal features. *Phonological Studies*, 12, 61–72.
- Jang, H. (2012). Acoustic properties and perceptual cues of Korean word-initial obstruents [국어 어두 장애음의 음향적 특성과 지각 단서]. Korea University dissertation.
- Jeon, L. (2011). *Drawing boundaries and revealing language attitudes: Mapping perceptions of dialects in Korea*. MA thesis, University of North Texas.
- Jin, W. (2008). Sounds of Chinese Korean: A variationist approach. University of Texas at Arlington dissertation.

- Jin, W., & Silva, D. (2017). Parallel Voice Onset Time shift in Chinese Korean. *Asia-Pacific Language Variation*, 3, 41–66. DOI: <https://doi.org/10.1075/aplv.3.1.03jin>
- Kang, Y. (2014). Voice Onset Time merger and development of tonal contrast in Seoul Korean stops: A corpus study. *Journal of Phonetics*, 45, 76–90. DOI: <https://doi.org/10.1016/j.wocn.2014.03.005>
- Kang, Y., & Han, S. (2012). Dialectal variation in Korean stops. *Proceedings of the 2012 Joint Meeting of the Society of Korean Linguistics and the Society of Korean Dialectology*.
- Kang, Y., Schertz, J., & Han, S. (forthcoming). The phonetics and phonology of Korean stop laryngeal contrasts. *The Cambridge Handbook of Korean Language and Linguistics*.
- Kim, M., Beddor, P., & Horrocks, J. (2002). The contribution of consonantal and vocalic information to the perception of Korean initial stops. *Journal of Phonetics*, 30, 77–100. DOI: <https://doi.org/10.1006/jpho.2001.0152>
- Kirby, J. (2013). The role of probabilistic enhancement in phonologization. *Origins of sound change: Approaches to phonologization*, Ed. by A. C. L. Yu (pp. 228–246). Oxford: Oxford University Press. DOI: <https://doi.org/10.1093/acprof:oso/9780199573745.003.0011>
- Kong, E., Beckman, M., & Edwards, J. (2011). Why are Korean tense stops acquired so early?: The role of acoustic properties. *Journal of Phonetics*, 39, 196–211. DOI: <https://doi.org/10.1016/j.wocn.2011.02.002>
- Koops, C., Gentry, E., & Pantos, A. (2008). The effect of perceived speaker age on the perception of PIN and PEN vowels in Houston, Texas. *U. Penn Working Papers in Linguistics*, 14.
- Lawrence, D. (2015). Limited evidence for social priming in the perception of the BATH and STRUT vowels. *Proceedings of the International Congress of Phonetic Sciences*.
- Lee, H. (2009). *Training program for the linguistic adaptation of new settlers*. The National Institute of the Korean Language: Seoul.
- 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, 145–169. DOI: <https://doi.org/10.1017/S0025100312000035>
- Lee, H., & Jongman, A. (2018). Effects of Sound Change on the Weighting of Acoustic Cues to the Three-Way Laryngeal Stop Contrast in Korean: Diachronic and Dialectal Comparisons. *Language and Speech*. Published online. DOI: <https://doi.org/10.1177/0023830918786305>
- 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, 117–132. DOI: <https://doi.org/10.1016/j.wocn.2012.12.002>
- Lisker, L., & Abramson, A. (1967). Some effects of context on Voice Onset Time in English stops. *Language and Speech*, 10(1), 1–28. DOI: <https://doi.org/10.1177/002383096701000101>
- Matuschek, H., Kliegl, R., Vasishth, S., Baayen, H., & Bates, D. (2017). Balancing Type I error and power in linear mixed models. *Journal of Memory and Language*, 94, 305–315. DOI: <https://doi.org/10.1016/j.jml.2017.01.001>
- Moulines, E., & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication*, 6, 453–467. DOI: [https://doi.org/10.1016/0167-6393\(90\)90021-Z](https://doi.org/10.1016/0167-6393(90)90021-Z)
- Niedzielski, N. (1999). The effect of social information on sociolinguistic variables. *Journal of Language and Social Psychology*, 18, 62–85. DOI: <https://doi.org/10.1177/0261927X99018001005>

- Oh, E. (2011). Effects of speaker gender on voice onset time in Korean stops. *Journal of Phonetics*, 39, 59–67. DOI: <https://doi.org/10.1016/j.wocn.2010.11.002>
- Oh, M., & Yang, H. (2013). The production of stops by Seoul and Yanbian Korean speakers. *Journal of the Korean Society of Speech Sciences*, 5, 185–193. DOI: <https://doi.org/10.13064/KSSS.2013.5.4.185>
- Ohala, J. (1993). Sound change as nature's speech perception experiment. *Speech Communication*, 13, 155–161. DOI: [https://doi.org/10.1016/0167-6393\(93\)90067-U](https://doi.org/10.1016/0167-6393(93)90067-U)
- Pharao, N., Lundholm Appel, K., Wolter, V., & Thogersen, J. (2015). Raising of /a/ in Copenhagen Danish – perceptual consequences across two generations. *Proceedings of the International Congress of Phonetic Sciences 2015*.
- Schertz, J., Cho, T., Lotto, A., & Warner, N. (2015). Individual difference in phonetic cue use in production and perception of a non-native sound contrast. *Journal of Phonetics*, 15, 183–204. DOI: <https://doi.org/10.1016/j.wocn.2015.07.003>
- Silva, D. (2006). Acoustic evidence for the emergence of tonal contrast in contemporary Korean. *Phonology*, 23, 287–308. DOI: <https://doi.org/10.1017/S0952675706000911>
- Squires, L. (2013). It don't go both ways: Limited bidirectionality in sociolinguistic perception. *Journal of Sociolinguistics*, 17, 200–237. DOI: <https://doi.org/10.1111/josl.12025>
- Staum Casasanto, L. (2010). What do listeners know about sociolinguistic variation? *U. Penn Working Papers in Linguistics*, 15.
- Strand, Elizabeth A., & Johnson, K. (1996). Gradient and visual speaker normalization in the perception of fricatives. In *KONVENS (Konferenz zur Verarbeitung Natürlicher Sprache)* (pp. 14–26). DOI: <https://doi.org/10.1515/9783110821895-003>
- Tai, P. (2004). Language policy and standardization of Korean in China. In *Language policy in the People's Republic of China: Theory and practice since 1949*, Ed. by Minglang Zhou & Hongkai Sun (pp. 303–315). Kluwer Academic Publishers. DOI: https://doi.org/10.1007/1-4020-8039-5_17
- Thomas, E. (2002). Sociophonetic applications of speech perception experiments. *American Speech*, 77(2), 115–147. DOI: <https://doi.org/10.1215/00031283-77-2-115>
- Willis, C. (1972). Perception of vowel phonemes in Fort Erie, Ontario, Canada, and Buffalo, New York: An application of synthetic vowel categorization tests to dialectology. *Journal of Speech, Language, and Hearing Research*, 15, 246–255. DOI: <https://doi.org/10.1044/jshr.1502.246>
- Yang, S. (2013). *Linguistics accommodation of North Korean refugees*. PhD thesis, Seoul National University.

How to cite this article: Schertz, J., Kang, Y., & Han, S. 2019 Sources of variability in phonetic perception: The joint influence of listener and talker characteristics on perception of the Korean stop contrast. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 10(1):13, pp. 1–32. DOI: <https://doi.org/10.5334/labphon.67>

Submitted: 19 December 2016 **Accepted:** 11 June 2019 **Published:** 16 July 2019

Copyright: © 2019 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.



Laboratory Phonology: Journal of the Association for Laboratory Phonology is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS