FEBRUARY 10 2021

Contrastive perception of duration and F0 cues for stop categorization[a)]

Sang-Im Lee-Kim

JASA Express Lett 1, 025203 (2021)
https://doi.org/10.1121/10.0003426

JASA EXPRESS LETTERS ARTICLE asa.scitation.org/journal/jel

Contrastive perception of duration and Fo cues for stop categorization[a)]

Sang-Im Lee-Kim
National Chiao Tung University, 1001 University Road, Hsinchu 300, Taiwan
sangimleekim@nctu.edu.tw

Abstract: This study examined the role of linguistic experience in the contrastive perception of suprasegmental cues for stop categorization. Korean-speaking learners of Mandarin and naive listeners labeled word-medial unaspirated stops (e.g., ma.**p**a) as either fortis [long closure–high F0] or lenis [short closure–low F0]. The results revealed comparable effects of relative duration for both groups: *shorter* neighboring vowels elicited more fortis responses, arising from *longer* perceived stop closure. However, F0 contours were processed contrastively only for the learners: stops were perceived as fortis before vowels with *lower* offset F0, which may have contributed to a *higher* perceived onset F0.

[Editor: Rachel M. Theodore] https://doi.org/10.1121/10.0003426

Received: 10 September 2020 **Accepted:** 12 January 2021 **Published Online:** 10 February 2021

1. Introduction

The perception of speech sounds arises from the integration of the intrinsic properties of target sounds and the extrinsic properties of neighboring sounds. The effect of extrinsic context is said to be "contrastive." For example, the processing of lexical tone is context-dependent for speakers of tone languages. Cantonese speakers were reported to identify an ambiguous level tone as *high* (/ji^{55}/ "doctor") in a context where the F0 of the sentence was lowered (*low*-F0 context), while the physically identical tone was identified as *low* (/ji^{22}/ "two") in a context where the F0 of the sentence was raised (*high*-F0 context) (Francis *et al.*, 2006). In a similar vein, the duration of stop closure is assessed in relation to the duration of a neighboring vowel. English listeners, for instance, perceive a stop segment as "voiceless," which is characterized by a *long* stop closure, when the preceding vowel has been *shortened* [e.g., Kluender *et al.* (1988) and Raphael (1972)]. Similarly, Japanese listeners are biased toward perceiving "geminates," or *long* consonants, when the preceding mora has been *shortened* (e.g., Idemaru and Guion-Anderson, 2010).

Contrastive perception is further modulated by linguistic experience. Kang *et al.* (2016), for example, reported that both English and French native speakers were more likely to categorize ambiguous noise signals as /s/ before /u/ than before /a/. The results indicate that the low F2 of the neighboring [u] caused the listeners to perceive the spectrally *higher* fricative [s]. However, such a perceptual contrast for the /y/ vowel, despite its relatively *high* F2, was only observed for French-speaking listeners, while English-speaking listeners treated /y/ and /a/ similarly. The results highlight the importance of language-specific linguistic knowledge of the sounds: lip rounding ordinarily involved in both /u/ and /y/ in French may have caused the two sounds to be grouped into a similar category, eliciting a robust perceptual contrast with neighboring fricatives.

Building upon this line of research, the present study explores how experience in a second language may build perceptual contrasts along newly acquired phonetic cues. In particular, learners of a tonal second language may process a perceptual contrast along F0 cues for non-native stop identification. This inquiry was tested against how phonetic cues to the fortis-lenis distinction are processed by Korean listeners in a word-medial position. The two stops are primarily distinguished by closure duration, with fortis stops having considerably longer closure than lenis stops [M(fortis) = 140 ms vs M (lenis) = 56 ms; Han (1996), p. 119]. However, the two stops can also be characterized by F0: regardless of position, vowels following fortis stops have a high F0 at the onset, whereas the onset of those coming after lenis stops has a low F0 (Cho and Keating, 2001). Taken together, the fortis and lenis stops can be represented by the following cue pairings, respectively: [high F0–long closure] and [low F0–short closure].

Capitalizing on the phonetic properties of the two stops, the present study investigates whether stop categorization is driven by a contrastive perception of those cues. First, stop closure may be perceived in relation to the neighboring

[a)]Portions of this work were presented in the Linguistic Society of America 2019 Annual Meeting, New York, NY, USA, 3–6 January 2019 and the 17th International Conference on the Processing of East Asian Languages and the 9th Conference on Language, Discourse, and Cognition (ICPEAL17–CLDC9), Taipei, Taiwan, 19–20 October 2018.

vowel duration. Not only are there large differences in absolute closure duration between fortis and lenis stops [e.g., Han (1996) and Silva (1992)], but vowel duration and stop closure tend to covary inversely: fortis stops are preceded by vowels with short duration, whereas lenis are preceded by vowels with long duration (Oh and Kim, 2019). It is conceivable that the perception of stop closure in Korean arises from a temporal relativization with locally adjacent segments. This study explicitly tested listeners' categorization of stops preceding vowels with varying durations, which will help establish how Korean listeners' contrastive perception operates along durational cues.

Beyond duration, we also examine the possibility of the onset F0 cue being relativized to the offset F0 to inform stop categorization. In addition to the absolute F0 value at the vowel onset, listeners may employ the entire F0 contour in their interpretation of onset F0. In two previous studies, rising F0 contours elicited more lenis judgments than level contours by Korean-speaking learners of Mandarin Chinese (Lee-Kim, 2020a,b). In these studies, however, level and rising contours were taken from physically different tokens, and other phonetic cues were not fully controlled. The current study was thus designed to isolate the role of F0 trajectories in the perception of onset F0, all else being equal. To keep other phonetic cues identical, a level contour was taken as a baseline from which rising and falling contours were generated. The falling contour was also included to strengthen empirical evidence for the contrastive perception hypothesis; a *high* offset F0 (rising contour) may trigger a lower perceived onset F0, while a low offset F0 (falling) may lead to a higher onset F0 perception. Accordingly, the former is expected to yield more lenis judgments, and the latter more fortis judgments. Acknowledging that lexical tone experience may have an influence on the sensitivity to F0 contour, this study compared Korean-speaking learners of Mandarin Chinese and those without prior exposure to tone languages.

2. Method

2.1 Participants

Thirty-seven native speakers of Seoul Korean were recruited for the perception study. The participants included 18 experienced learners of Mandarin (12 female, 6 male; aged 20–24) and 20 naive listeners (12 female, 8 male; aged 20–28). The language background survey completed by the learners indicated that the mean length of Mandarin was 7.3 years [standard deviation (SD) = 3.9). Most learners had passed advanced (level 6, 11 learners) or intermediate levels (level 4 or 5, 7 learners) of the HSK test (a Chinese proficiency test). Nine learners reported having lived in China, from as short as 1 month to as long as 4 years. The naive listeners had no experience in Mandarin or any other tone languages and spoke Seoul Korean, which is neither tonal nor stressed (Jun, 2005). All participants had learned English in secondary school. None of the participants reported speech or listening impairments. All were monetarily compensated for their time.

2.2 Stimuli manipulation

The stimuli used in the experiment were made by digitally manipulating two disyllabic nonce words, /ma^{55}.pa^{55}/ and /na^{55}.ta^{55}/. A female native speaker of Mandarin produced the tokens in isolation three times while reading from a randomized list that included other filler items. The recordings were made using a Zoom H4n recorder connected to a Shure SM58 microphone and digitized at a sampling rate of 22,050 Hz in 16-bit quality.

Clearly produced tokens of each word were chosen and manipulated along three phonetic dimensions. First, the durations of both preceding and following vowels (V_1 and V_2, respectively) were varied between short (300 ms), medium (350 ms), and long (400 ms) using pitch-synchronous overlap and add (PSOLA) re-synthesis in Praat (Boersma and Weenink, 2017), while the intervening stop closure was set to be 90 ms, an unbiased value between fortis ($M = 140$) and lenis ($M = 56$) stop closure [Han (1996), p. 119]. Second, F0 contours were created by varying the offset F0 of V_2 between high, mid, and low by raising or lowering the offset of the original level contour by 60 Hz.[1] The onset and offset F0s were interpolated by the quadratic function in the PSOLA algorithm, resulting in different contours: rising (high offset), level (mid offset), and falling (low offset). In addition, the onset F0 of V_2, "F0 step," was varied by 20 Hz on a three-step continuum: 180–200–220 Hz. V_1 always carried a level tone of the same F0 as the onset of V_2. This variable was motivated to obtain the baseline performance against the F0 cue for the fortis-lenis distinction between the two groups of participants. Varying F0 contours were created for each F0 step. Figure 1 illustrates an example of manipulated tokens.

The particular acoustic values were chosen to ensure that the stimuli would sound natural after digital manipulation. The F0 range of the female talker in this study was slightly higher than reported averages (Peng *et al.*, 2012), ranging from 150 Hz (onset of tone 2) to 360 Hz (onset of tone 4). The raised F0 offset for the rising contour fell well within the talker's normal F0 range [e.g., 220 Hz (onset F0) ± 60 Hz = 280 Hz (offset F0)]. The lowered F0 offset resulted in a slightly lower F0 value than the talker's F0 range [e.g., 180 Hz (onset F0) − 60 Hz = 120 Hz (offset F0)], but no acoustic distortion was apparent in the signal. Likewise, the duration values (300–350–400 ms) were chosen based on the talker's vowel durations, which generally fell between 300 and 400 ms to minimize the extent of digital manipulation.[2]

2.3 Procedure

The perception experiment was run in E-PRIME in a sound-attenuated booth at the Phonetics Laboratories at Seoul National University. Participants wore AKG-K240 headphones connected to a laptop computer and were tested individually. The stimuli set consisted of five blocks, each containing 54 stimuli (3 VDs ∗ 3 F0 contours ∗ 3 F0 steps ∗ 2 words),

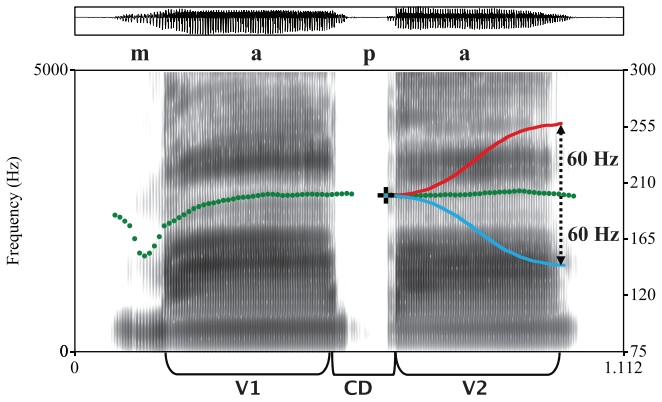

Fig. 1. An example spectrogram of the stimulus /mapa/. Vowel durations (VDs: $V_1 = V_2$) were manipulated to be 300, 350, and 400 ms, while closure duration (CD) was fixed at 90 ms. Given an onset F0 value indicated by a cross symbol (200 Hz in this figure), the F0 offset of a level contour was raised by 60 Hz, unchanged, or lowered by 60 Hz to form different F0 contours: rising (red), level (green), and falling (blue).

which were randomized within each block. In a two-alternative forced-choice identification task, participants were asked to label a stimulus either as lenis or fortis using the number keys on the keyboard as quickly and accurately as possible. The two options were given in Korean orthography alongside the number keys, for example, (1) 마바 ([ma**p**a] for lenis) vs (2) 마빠 ([ma**p**'a] for fortis). The order of the two stops was fixed throughout the experiment, with lenis on the top and fortis on the bottom. Before the test session, five practice trials were provided to familiarize the participants with the procedure. The experiment lasted about 15 min.

A mixed-effects logistic regression model was fitted to the data using the *glmer* function in the *lme4* package (Bates *et al.*, 2015) implemented in R 3.2.2 (R Development Core Team, 2020). The dependent variable was the participants' binary judgment of the stop stimuli: fortis (coded as "1") or lenis (coded as "0"). Fixed effects for the model included DURATION (3 levels: short, medium, long), CONTOUR (3 levels: rise, level, fall), F0STEP (3 levels centered: 180–200–220), and GROUP (2 levels contrast-coded: learners = −1 vs naive = 1). Interaction terms between GROUP and each fixed variable were included to examine whether the two groups of listeners responded to the variables differently. Additionally, the CONTOUR–DURATION interaction was initially considered since the F0 rate (F0 change/duration) could potentially affect contour perception; however, the larger model did not converge, so it was included only in the *post hoc* analyses. The original model included random intercepts for SUBJECT and WORD as well as random slopes of DURATION and F0STEP for SUBJECT. The model failed to converge with CONTOUR as a random slope for SUBJECT, and it was therefore included in the *post hoc* analyses conducted for each group's data to keep the focus on the contour-driven differences in stop categorization.

3. Results

Figure 2 plots the estimated response curves drawn from the regression model of the participants' stop judgments as a function of F0 step. The results from the different vowel durations are summarized in different columns for learners (top) and naive listeners (bottom). Responses to different contours are represented in different colors. Table 1 summarizes the results of the statistical analyses.

Figure 2 demonstrates that for both groups, more lenis judgments were given as vowel duration increased. The results of the regression analysis confirmed this observation. For the level contours, vowels with medium durations elicited more lenis responses than did those with short durations ($\beta = 0.5356$, $p < 0.0001$) but fewer lenis responses than those with long durations ($\beta = -0.5097$, $p < 0.0001$). The insignificant interactions between DURATION and GROUP indicate this pattern held true for both groups. Taken together, the results suggest that the perception of stop closure arises from a temporal relativization to the neighboring vowels. That is, when the neighboring vowels are *long*, the stop is perceived as having shorter closure, which is a characteristic of lenis stops.

The variable F0STEP was motivated to obtain the baseline sensitivity to F0 cues by the two groups of listeners. In the baseline condition, the statistical analysis revealed that F0STEP is a strong predictor of listeners' stop judgments; namely, higher F0 prompts more fortis responses ($\beta = 0.5867$, $p < 0.0001$). Regardless of experience, the two groups were comparable with respect to this variable as indicated by the non-significant F0STEP–GROUP interaction ($\beta = -0.0153$, $p = 0.782$). This pattern can be visually confirmed via the positive slopes for the vowels with medium durations and level contours in Fig. 2. This result indicates that F0 is indeed a legitimate phonetic cue for the distinction of word-medial fortis-lenis stops.

Despite the comparable sensitivity to F0STEP by both groups of Korean listeners, the two groups responded differently to the various F0 contours as indicated by the highly significant CONTOUR–GROUP interactions ($CONTOUR_{fall}$:

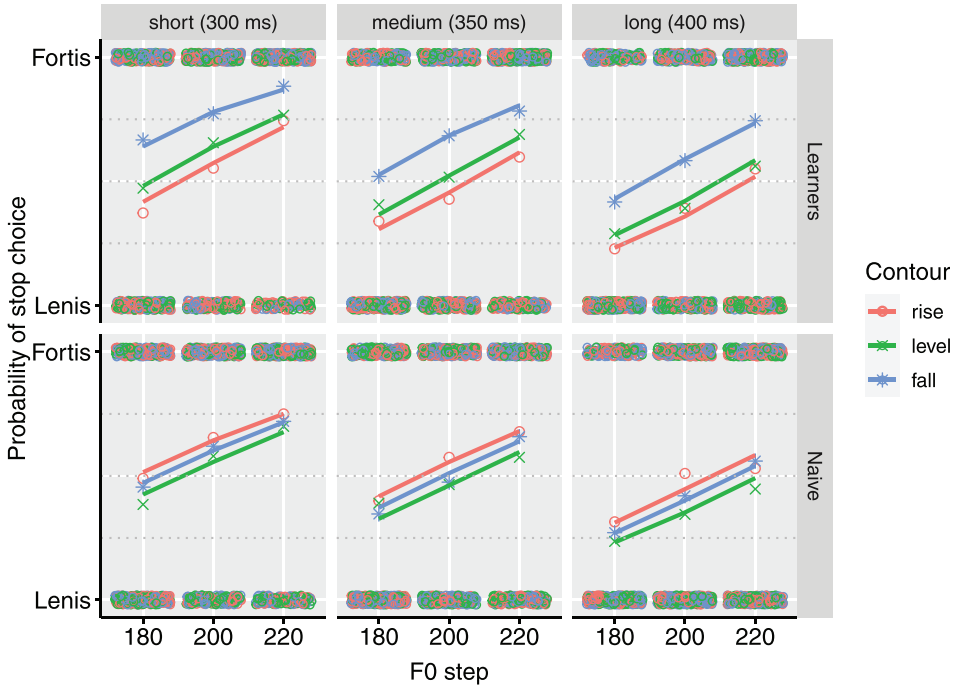

Fig. 2. Predicted logit curves and actual mean values for stop identification plotted against F0 step and vowel duration (different columns). Results from the learners and naive listeners are summarized in the top and bottom panels, respectively.

GROUP: $\beta = -0.2487$, $p < 0.0001$; $\text{CONTOUR}_{\text{rise}}$:GROUP: $\beta = -0.3863$, $p < 0.0001$). To determine the source of these interactions, separate regression models were fitted to the data from each group.[3] The analysis of the data from the learners with medium vowel duration set as the baseline revealed that the falling contour elicited significantly more fortis judgments than the level contour ($\beta = 0.8595$, $p = 0.021$), which in turn elicited marginally more fortis judgments than the rising contour ($\beta = -0.4542$, $p = 0.058$). As suggested by the insignificant interaction between CONTOUR and DURATION (p values > 0.100), this contour-dependent perceptual pattern generally held true across all durations. From the relative positioning of the contours in Fig. 2, we can see that the falling contour (blue) yielded the most fortis judgments, the rising contour (red) the least, and the level contour (green) fell in between.

As for the naive listeners, results from the statistical model revealed an entirely different picture. With medium vowel duration and level contour set as the baseline, a minor trend was observed such that the rising contour led to more fortis judgments than the level contour ($\beta = 0.3794$, $p = 0.082$) without further interaction with DURATION. This reflects the few but consistently increased fortis judgments for the rising contour across vowel durations observed in Fig. 2. The falling contour was not significantly different from the level contour in the baseline condition ($\beta = 0.0935$, $p = 0.629$).

Table 1. Summary of the mixed-effects logistic regression model predicting stop categorization. The values for "medium" duration and "level" contour serve as reference levels. Significant results are represented in bold. SE, standard error.

Predictor	Coefficient	SE	z value	p value
(Intercept)	−0.0520	0.1789	−0.290	0.771
GROUP	−0.1705	0.1598	−1.067	0.286
$\text{DURATION}_{\text{short}}$	**0.5356**	**0.0741**	**7.226**	**<0.0001*****
$\text{DURATION}_{\text{long}}$	**−0.5097**	**0.0622**	**−8.193**	**<0.0001*****
$\text{DURATION}_{\text{short}}$:GROUP	−0.0234	0.0736	−0.318	0.751
$\text{DURATION}_{\text{long}}$:GROUP	−0.0538	0.0620	−0.868	0.385
F0STEP	**0.5867**	**0.0553**	**10.601**	**<0.0001*****
F0STEP:GROUP	−0.0153	0.0553	−0.277	0.782
$\text{CONTOUR}_{\text{fall}}$	**0.4954**	**0.0560**	**8.842**	**<0.0001*****
$\text{CONTOUR}_{\text{rise}}$	0.0865	0.0553	1.563	0.118
$\text{CONTOUR}_{\text{fall}}$:GROUP	**−0.2487**	**0.0560**	**−4.440**	**<0.0001*****
$\text{CONTOUR}_{\text{rise}}$:GROUP	**0.3863**	**0.0553**	**6.984**	**<0.0001*****

However, the small p value of the interaction between $\text{DURATION}_{\text{long}}$ and $\text{CONTOUR}_{\text{falling}}$ ($\beta = 0.3426$, $p = 0.075$) indicates that the contour-driven differences became larger as vowel duration increased.[4] Bearing in mind the small effect size, however, naive listeners were, overall, unaffected by F0 later into the vowel, and F0 contour was inconsequential to their stop judgments.

In summary, group differences were unambiguous. For the learners, fortis judgements followed a clear pattern depending on contour type: falling > level > rising. In contrast, a weak trend with a different order was observed for naive listeners: rising > level and falling ≧ level. Crucially, the learners perceived the stop before the rising contour as lenis, whereas the naive listeners were more likely to perceive the same stop as fortis.

4. Discussion

The present study examined whether the phonetic cues to the lenis-fortis distinction in Korean are processed relative to the phonetic attributes of the surrounding vowels. In Korean, fortis stops are associated with high F0 and long closure, whereas lenis stops are characterized by a low F0 and short closure. A perception experiment examined if stops with a neutral closure duration would be perceived differently depending on the duration and tonal contour of the neighboring vowels. To determine whether contrastive perception developed with linguistic experience, we included both experienced learners of a tone language as well as naive listeners.

The effect of vowel duration on the perceived closure duration was straightforward. The longer the vowel duration, the more likely the listeners in both groups were to perceive lenis stops. Fortis stops, on the other hand, were perceived when the neighboring vowel had a longer duration. That is, *longer* vowel duration leads to *shorter* perceived stop closure, and vice versa. This pattern suggests that the absolute duration of stop closure is not sufficient; rather, listeners utilize the temporal relationship between adjacent segments for stop categorization. This may also be a reflection of the Korean listeners' phonetic knowledge of the covariance between vowels and stop closure in their native language (Oh and Kim, 2019).

In addition to vowel duration, we found that F0 contour plays a role in contrastive perception for those with experience in Mandarin. The target stops were positioned before vowels carrying three different tone contours: falling (*low* offset F0), level (*mid* offset F0), and rising (*high* offset F0). With the high offset F0, learners tended to perceive the preceding stops as lenis. This result was taken to reflect the contribution of offset F0 to the perceived onset F0: the onset F0 of a rising contour may be perceived as *lower* due to the perceptual contrast with the *high*-F0 offset. Conversely, the onset F0 of a falling contour may be perceived as *higher* due to the contrast with the *low*-F0 offset. Moreover, the level contour, with an F0 offset between that of the falling and rising contours, yielded an intermediate number of fortis and lenis responses, providing supporting evidence that the offset F0 is a good predictor of stop categorization by the learners.

Nevertheless, it is worth noting that the contrast between onset and offset F0 is not a typical case of contrastive perception, which usually involves clear-cut boundaries between a target segment and a separate context segment. For example, target and context tones are carried by different syllables in lexical tone languages [e.g., Francis *et al.* (2006)], and target stop closure and context vowel duration are temporal measures of two distinct segments [e.g., Idemaru and Guion-Anderson (2010) and Kluender *et al.* (1988)]. In the case of F0 contour, however, the change in F0 unfolds gradually throughout the duration of a single vowel, and it is not readily understood how the perceptual contrast between onset and offset F0 arises. We conjecture that the perceptual contrast along the gradually changing F0 may reflect the nature of speech perception. Speech perception develops through learning to categorize continuous speech signals into discrete units. The Korean learners of Mandarin may have developed the ability to decompose continuous F0-trajectories into categorical subcomponents. As proposed in the phonological literature (Flemming and Cho, 2017), the rising contour, for example, may be broken down into two tonal targets, namely LH, which may enable contrastive perception between the two distinct F0 targets.

In contrast to those with experience in lexical tones, naive Korean listeners were, overall, less sensitive to the F0 contour variable, as evidenced by the ostensibly small differences in their responses to the different contours. Recall that the naive listeners were able to utilize F0 step to the same degree as the learners, as shown in the positive slopes in Fig. 2. This suggests that naive listeners were sensitive to the onset F0 and actively incorporated this cue into stop categorization. That being said, the F0 trajectory, especially the F0 later into the vowel, was, by and large, irrelevant to the stop judgments for those listeners.

Taken together, the findings of the current study suggest that patterns of stop identification for the learners develop over time with experience: experience with dynamic F0 contours may enable learners to employ an additional F0 contrast beyond a simple temporal contrast in stop categorization.

Acknowledgments

The author would like to thank Hyoju Kim and Sarang Jeong for their help with data collection and Hannah Parsamehr for her careful editing of the manuscript. The author also thank audiences at the Linguistic Society of America (LSA 2019) and the 17th International Conference on the Processing of East Asian Languages and the 9th Conference on Language, Discourse, and Cognition 2018 (ICPEAL17–CLDC9). The reviewers of JASA Express Letters provided constructive comments, which led to

substantial improvements to the manuscript. This work was supported by Research Grant No. MOST107-2410-H-009-016-MY3 from the Ministry of Science and Technology, Taiwan.

References and links

[1]It is not clear how much of a fall or rise in F0 is sufficient for the perception of contour tones. This particular value was deemed reasonable as the contours were auditorily salient while still sounding adequately natural.

[2]These particular vowel durations are presumably longer than those reported for Korean vowels (Oh and Kim, 2019). The longer neighboring vowel durations could have biased Korean listeners toward perceiving lenis (short closure duration) over fortis (long closure duration). On the other hand, the lack of voicing during closure could have biased listeners toward perceiving fortis (no voicing) over lenis stops (some voicing) (Silva, 1992). The results in Fig. 2 show no evidence of perceptual biases toward either stops, suggesting that these opposing acoustic cues to the laryngeal contrasts may have negated each other.

[3]All R codes and statistical summary tables are available at https://osf.io/24syq/.

[4]The same baselines (medium vowel duration, level contour) were used across different statistical models for the sake of consistency, but when long vowel duration was set as the baseline, the results showed unambiguously significant effects of contour ($CONTOUR_{fall}$: $\beta = 0.4103$, $p = 0.0003$; $CONTOUR_{rise}$: $\beta = 0.5900$, $p = 0.0001$). The minor trend observed in the medium vowel duration was amplified and verified in long vowel duration, suggesting that the listeners without experience with tone languages may be able to process F0 contours only when vowel durations are sufficiently long. It is not entirely clear why the naive listeners perceived fortis stops more often when the vowel carried contour tones (i.e., rising and falling) than when the tone was level. One possible explanation is that the naive Korean listeners were less sensitive to the direction of the F0 change over time but were able to pick up on the high F0 component during the vowel, which may have biased them toward categorizing the stop as fortis.

Bates, D., Maechler, M., Bolker, B., and Walker, S. (**2015**). "Fitting linear mixed-effects models using lme4," J. Stat. Softw. **67**(1), 1–48.

Boersma, P., and Weenink, D. (**2017**). "Praat: Doing phonetics by computer (version 6.0.26) [computer program]," www.praat.org (Last viewed 12 December 2018).

Cho, T., and Keating, P. A. (**2001**). "Articulatory and acoustic studies on domain-initial strengthening in Korean," J. Phon. **29**, 155–190.

Flemming, E., and Cho, H. (**2017**). "The phonetic specification of contour tones: Evidence from the Mandarin rising tone," Phonology **34**, 1–40.

Francis, A. L., Ciocca, V., Wong, N. K. Y., Leung, W. H. Y., and Chu, P. C. Y. (**2006**). "Extrinsic context affects perceptual normalization of lexical tone," J. Acoust. Soc. Am. **119**, 1712–1726.

Han, J.-I. (**1996**). "The phonetics and phonology of 'tense' and 'plain' consonants in Korean," Ph.D. thesis, Cornell University, Ithaca, NY.

Idemaru, K., and Guion-Anderson, S. (**2010**). "Relational timing in the production and perception of Japanese singleton and geminate stops," Phonetica **67**, 25–46.

Jun, S.-A. (**2005**). "Korean intonational phonology and prosodic transcription," in *Prosodic Typology: The Phonology of Intonation and Phrasing* (Oxford University Press, Oxford, UK), pp. 201–229.

Kang, S., Johnson, K., and Finley, G. (**2016**). "Effects of native language on compensation for coarticulation," Speech Commun. **77**, 84–100.

Kluender, K., Diehl, R. L., and Wright, B. (**1988**). "Vowel-length differences before voiced and voiceless consonants: An auditory explanation," J. Phon. **16**, 153–169.

Lee-Kim, S.-I. (**2020a**). "Stop laryngeal distinctions driven by contrastive effects of neighboring tones," Lang. Speech (published online).

Lee-Kim, S.-I. (**2020b**). "Tone-induced split in stop category mapping by Korean learners of Mandarin Chinese," Lang. Linguist. **21**(1), 54–79.

Oh, M., and Kim, D. (**2019**). "Shortening of voice onset time in Korean aspirates: Phonetic OCP or timing regulation?," in *Proceedings of the 19th International Congress of Phonetic Sciences*, August 5–9, Melbourne, Australia, pp. 657–660.

Peng, G., Zhang, C., Zheng, H.-Y., Minett, J. W., and Wang, W. S.-Y. (**2012**). "The effect of intertalker variations on acoustic–perceptual mapping in Cantonese and Mandarin tone systems," J. Speech Lang. Hear. Res. **55**, 579–595.

Raphael, L. J. (**1972**). "Preceding vowel duration as a cue to the perception of the voicing characteristics of word-final consonants in English," J. Acoust. Soc. Am. **51**, 1296–1303.

R Development Core Team (**2020**). "R: A language and environment for statistical computing," R Foundation for Statistical Computing, Vienna, Austria, http://www.r-project.org/ (Last viewed 15 August 2020).

Silva, D. J. (**1992**). "The phonetics and phonology of stop lenition in Korean," Ph.D. thesis, Cornell University, Ithaca, NY.