

Perceptual integration of acoustic cues to laryngeal contrasts in Korean fricatives

Sarah Lee, Dept. of linguistics, University of California Berkeley, Dwinelle Hall, Berkeley,
California 94720

Jonah Katz^a, Department of world languages, literatures, & linguistics, West Virginia University,
Chitwood Hall, P.O. Box 6298, Morgantown, West Virginia 26506-6298

DATE

Running title: Cue integration and Korean fricatives

^a Author to whom correspondence should be addressed. Electronic mail: katzlinguist@gmail.com

Abstract

This paper provides evidence that multiple acoustic cues involving the presence of low-frequency energy integrate in the perception of Korean coronal fricatives. This finding helps explain a surprising asymmetry between the production and perception of these fricatives found in previous studies: lower F0 onset in the following vowel leads to a response bias for plain [s] over fortis [s*], despite the fact that there is no evidence for a corresponding acoustic asymmetry in the production of [s] and [s*]. A fixed classification task using the Garner paradigm provides evidence that low F0 in a following vowel and the presence of voicing during frication perceptually integrate. This suggests that Korean listeners in previous experiments were responding to an ‘intermediate perceptual property’ of stimuli, despite the fact that the individual acoustic components of that property are not all present in typical Korean fricative productions. The finding also broadens empirical support for the general idea of perceptual integration to a new language, a different manner of consonant, and a situation where covariance of the acoustic cues under investigation is not generally present in a listener’s linguistic input.

PACS numbers: 43.71.Es, 43.71.An

I. INTRODUCTION

Most phonological contrasts involve a number of independent phonetic differences. For instance, the difference between /m/ and /n/ is cued by lower second and third resonances for /n/ before most vowels (Kurowski & Blumstein 1984), downward-sloping F2 transitions from a preceding vowel into /m/ (Liberman *et al.* 1954), and upward-sloping F2 transitions from /m/ into a following vowel (Malecot 1956); all three sets of differences make some contribution to place perception (Malecot 1956). The question of how a listener integrates multiple acoustic cues such as these is a foundational issue in phonetics and speech processing. This paper provides evidence from the perception of Korean fricatives that some kinds of cue integration are driven by low-level auditory properties and do not require linguistic experience to be learned.

Such cases of cue integration are interesting because prior research has sometimes argued that the integration process fundamentally relies on linguistic experience: listeners learn that certain acoustic cues tend to covary by hearing speech where they covary, and consequently they are able to bind multiple cues together into a coherent linguistic percept (Kluender, 1994; Nearey, 1997). This type of theory, which we label *empiricist integration* following Nearey, is compelling and empirically well-supported. It is difficult to extend, however, to cases where the perception of speech sounds depends on acoustic properties that are not consistently present in their productions. This paper explores a particular case of this type and proposes that it can be explained with the help of another kind of theory, which we label *auditory integration*, following Kingston *et al.* (2008). In this view, certain acoustic cues are inherently more likely to integrate in speech perception because they have similar effects on the human auditory system (Parker, Diehl, & Kluender, 1986; Kingston & Diehl, 1994; Kingston *et al.*, 2008). Note that these two types of theory are by no means mutually exclusive. Given that the human auditory system treats

some acoustic properties differently from others, and given that acoustic covariation is pervasive in language, it is perfectly sensible for both mechanisms to play a role in speech perception.

The asymmetry investigated here involves Korean coronal fricatives that contrast for laryngeal specifications. While the complex 3-way contrast amongst Korean stops has attracted a lot of attention in the phonetic literature (e.g. Han & Weitzman, 1970; Abramson & Lisker, 1973; Dart, 1987; Silverman & Jun, 1994; Cho, Jun & Ladefoged, 2002), only a fraction of these studies address the properties of the 2-way contrast for laryngeal features in coronal fricatives. The two fricative phones are referred to here as *fortis* [s*] and *non-fortis* [s]. Several acoustic properties distinguish the two sounds from one another (Yoon, 1999; Cho, Jun, & Ladefoged, 2002; Kim *et al.*, 2010). Fortis [s*] involves glottalization, which in turn affects the voice quality of a following vowel, and it lacks aspiration in all contexts. Non-fortis [s] induces breathy voice in the following vowel; it is aspirated word-initially and unaspirated medially. Medial [s] displays variation in the presence and extent of voicing during frication, up to and including tokens with voicing throughout (Cho, Jun, & Ladefoged, 2002). Chang (2013) presents a comprehensive review of these and other differences between the two fricatives. Here, we focus on differences related to voicing and the F0 onset of a following vowel.

This contrast was chosen because previous literature suggests an interesting disconnect between the production and perception of the sounds involved. Cho, Jun, & Ladefoged (2002) and Chang (2013) find no significant differences between the two fricatives with regard to the F0 onset of a following vowel. This holds for both initial and medial tokens.¹ Chang's perception experiment, however, finds that listeners consistently identify ambiguous tokens as fortis [s*] more often when the F0 onset of the following vowel is higher. This is a somewhat surprising

¹ Cho, Jun, & Ladefoged (2002) did find a significant F0 difference in the Cheju variety, but not in the Seoul variety investigated here and in Chang's work.

result: if fortis fricatives are not produced with higher F0 onset on a following vowel, why are listeners more likely to identify fricatives as fortis when the following F0 onset is higher?

We propose here that a low F0 onset following non-fortis fricatives perceptually integrates with other cues that *are* reliably present in production, because they have similar auditory effects. In particular, lower F0 in the following vowel integrates with at least the presence of voicing during frication, which is sometimes but not always present in domain-medial non-fortis fricatives. Both of these cues serve to increase the amount of low-frequency energy in the acoustic signal in the vicinity of frication, a property that sets the non-fortis fricative apart from its glottalized counterpart. For this reason, low F0 biases listeners towards non-fortis responses even though it is not reliably present in production of this segment.

The idea that low-frequency energy is an important ‘intermediate perceptual property’ for laryngeal contrasts is due to Kingston *et al.* (2008). In a series of experiments, they show that a low onset of both F0 and F1 in a following vowel integrate with the presence of closure voicing when English speakers classify synthetic vowel-stop-vowel stimuli, as well as non-speech analogues. Their conclusion is that these cues integrate at a low level of auditory perception, as shown by the non-speech analogues. They also conclude, based on the fact that neither low F0 nor low F1 onset integrate with closure duration, that experience of correlation in language is not sufficient to drive integration.

Kingston *et al.* (2008) use Garner’s (1974) paradigm for their experiments, and we adopt this paradigm as well. The idea is that, if two acoustic dimensions are perceptually independent, then covarying them will have symmetrical consequences for discrimination; if they are (partially) integrated, however, covarying the two dimensions in an *integrative* way will result in a more perceptible contrast than covarying them in an *oppositional* way. As an example, assume

that low F1 and low F0 perceptually integrate. A vowel with low F0 and F1 will be particularly distinct from a vowel with high F0 and F1, because the two cues ‘work together’ in the contrast (hence the label *integrative*). A vowel with low F0 and high F1, on the other hand, will not be as distinct from a vowel with high F0 and low F1, because here the two cues are working ‘at cross purposes’ (hence the label *oppositional*).

To test whether two dimensions integrate perceptually, then, we need to covary the two dimensions in both directions, and ask whether integrative covariance results in contrasts more perceptible than oppositional covariance. This is illustrated schematically in figure 1. We probe perceptual distance here using experimental measures of sensitivity.

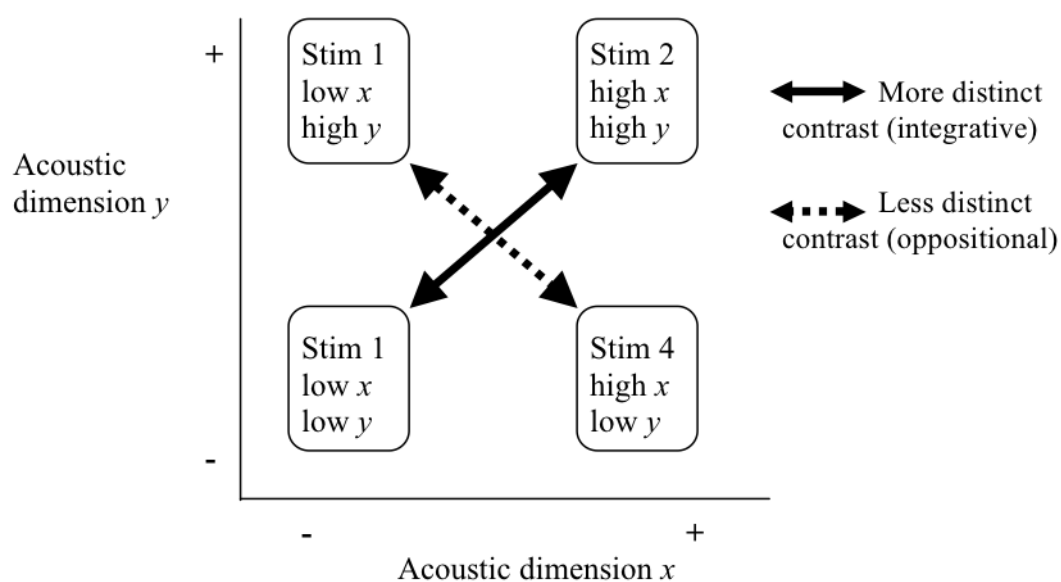


FIG 1. Schematic illustration of two acoustic dimensions, x and y , that integrate during the course of perception. The contrast between the stimulus where both values are high and the one where both values are low is *integrative*: the cues work together and the contrast is more perceptible. The contrast between the stimuli where one value is high and one low is *oppositional*: the cues work against each other and the contrast is less perceptible.

In what follows, we test whether low following F0 onset and voicing during frication integrate for Korean coronal fricatives. The test involves comparing discrimination between stimuli that differ in a putatively integrative way and stimuli that differ in a putatively oppositional way. There are two broad reasons why the question is interesting. First, if the cues integrate, it entails that they have some degree of perceptual equivalence, which would help explain why Korean listeners use the F0 cue for identifying laryngeal features of these fricatives even though that cue is not reliably present in production. Second, the experiment attempts to replicate and extend some of the findings of Kingston *et al.* (2008) regarding the low-frequency property. Their hypothesis that integration happens at a low level of audition, with limited or no interference from linguistic knowledge, suggests that the results should generalize quite widely across languages. And Korean offers a somewhat different and interesting test of the idea that the integration does not rely on linguistic experience: Korean speakers ostensibly have no experience with low F0 onsets correlating with voicing during fricatives, so if these cues integrate it cannot be explained straightforwardly as a consequence of prior linguistic exposure.

II. METHODS

A. Stimuli

We constructed a Garner paradigm (Garner 1974) for the dimensions of voicing during frication and F0 onset of a following vowel, using altered stimuli based on natural variation in the production of intervocalic non-fortis fricatives in Seoul Korean. The set-up allows us to test integration of cues through discrimination *within* a phonological category, rather than between categories. This is desirable because we are investigating an effect that is putatively not

influenced by linguistic knowledge, which would be more difficult if the task involved phonological categorization.

We tested the discriminability of each of the two acoustic cues independently (we refer to these conditions as ‘simple contrast’), then combined the two cues in both an oppositional and an integrative way. Comparing the oppositional contrast condition to the simple contrast conditions will test whether oppositional covariation reduces discriminability. Comparing the integrative contrast to the oppositional one will test whether the two cues integrate perceptually.

A female native speaker of Seoul Korean was recorded reading four tokens of several words with fortis and non-fortis fricatives in a variety of carrier sentences. The utterances were recorded with a condenser microphone in a sound-attenuated booth in the UC Berkeley Phonology Laboratory, using the Praat software (Boersma & Weenink n.d.). We selected a natural token of the word [kisuks*a] ‘dormitory’ as the base for our stimuli, with the first, non-fortis fricative being the focus. Out of the non-fortis fricatives we recorded, the one in this word was most frequently realized with voicing throughout the consonant; recall that intervocalic voicing of this sound is gradient and optional. We selected a voiced token because editing a stimulus to remove voicing is generally easier and more natural-sounding than editing it to add voicing. There was substantial variability in the F0 onset following the non-fortis fricative in recorded tokens of this word, which confirms that it is feasible to vary this parameter while still remaining within category boundaries of natural speech.

As the base token was voiced throughout the target fricative, its voicing properties were left unchanged for the voiced stimuli. For partially devoiced stimuli, voicing was retained for the first quarter of frication duration and removed from the remaining three quarters. Initial voicing was retained because a short interval of voicing at the beginning of the fricative was the most

common variant observed in the materials we recorded. To create devoiced stimuli, we used the pass Hann band filter function in Praat to remove all energy below 1000 Hz. This eliminated F0 and the first several harmonics (which created the percept of a fundamental if not removed). We defined frication duration as beginning and ending at points where the amount of energy above 5 kHz changed suddenly in the spectrogram; this tended to include portions of what might otherwise be considered the preceding and following segmental transitions.

The voiced and devoiced tokens resulting from this initial manipulation were then altered with regard to their F0 contours. We used Praat to create manipulation objects with 5 ms windows, then extracted pitch tiers from them. The pitch tiers were manually altered to raise or lower the original F0 onset following frication (250 Hz. in the original token) by 30 Hz, which fell within the natural range of variation attested in the recordings: vowel-onset F0 frequencies were thus 220 Hz for low stimuli and 280 Hz for high stimuli. Subjects were not exposed to stimuli with the original 250 Hz onset. The three pitch points following the one at the annotated segment boundary (which was raised or lowered 30 Hz) were also altered to create a smooth transition to the fifth point. The result is that the high and low stimuli differ in the first 25 ms of the vowel following the target fricative (about 45% of this rather short vowel's duration), with a difference of 60 Hz. at the onset and successively smaller differences at each following time step. In addition, we changed the contour at the end of frication for the fully-voiced stimulus with low F0 onset, in order to avoid a sequence of rapid F0 reversals; this difference is visible in the figure 2 below, which shows the four stimuli synthesized for the experiment.

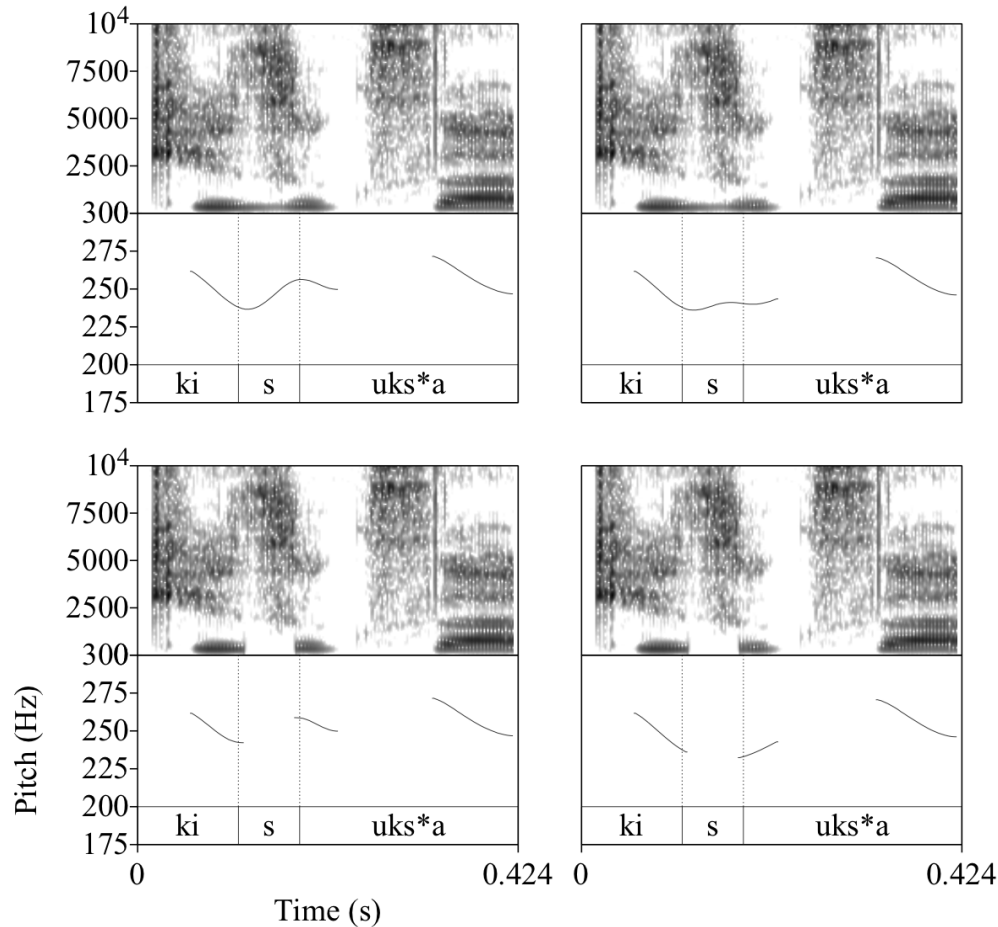


FIG. 2. Spectrograms and (smoothed) F0 contours of the four stimuli used in the experiment. The two in the top row are voiced, the two in the bottom row devoiced. The two on the left have higher F0 onset at the end of frication, the two on the right lower F0 onset.

Previous research suggests that just-noticeable-differences (JNDs) in F0 and/or F0 movement for a wide variety of level and contour tones are much smaller than the 60 Hz (4.18 st in this F0 range) used here (Flanagan & Saslow 1958, Klatt 1973, t'Hart 1981, Liu 2013). We used larger values because the differences here are very short in duration, about 25 ms. Impressionistically, the stimuli in all conditions were difficult to discriminate. The simple F0 contrasts, in particular, were very hard to hear. We ran the simple-contrast conditions on two pilot subjects to test

whether the task was feasible; both subjects performed just slightly above chance (31-36 correct out of 60 in each block).

B. Experimental design and procedure

The experimental procedure followed Kingston *et al.* (2008) wherever possible, so as to ensure comparability of results. It consisted of 6 blocks, each featuring two of the four stimuli. The order of blocks was separately randomized for each subject. The blocks separately examined the voiced/devoiced contrast at low F0 onset and high F0 onset, the F0 onset contrast for voiced and devoiced stimuli, and the two contrasts that vary both F0 onset and voicing. In terms of figure 1, these conditions can be thought of as the four sides of the square defined by the stimuli (simple contrasts) and the two diagonals of that square (oppositional and integrative contrast).

Each block consisted of 20 randomized training trials with feedback and 60 randomized test trials with feedback (30 responses per stimulus per block), without a gap in between. In each trial, the subject heard one of the stimuli and had an unlimited time to classify it, after which a two-second feedback screen would appear ('correct' or 'incorrect'). The blocks were presented in random order; in between blocks, subjects were given the option to take a break and press a button when they were ready to continue.

Subjects were told that there were two different sounds in each block, 'A' and 'B'; that they would learn what the two sounds were through trial and error at the beginning of the block; that the sounds would change with each block; and that the sounds would be difficult to tell apart. They were asked to label the sound played as either 'A' or 'B'.

The experiment was run in a quiet room and no more than two subjects were run at a time. The experiment was designed and run with the E-Prime 2.0 software. Stimuli were played

through AKG K240 semi-open studio headphones. Responses were recorded on a standard computer keyboard.

C. Subjects

14 native Korean speakers participated in the experiment, seven male and seven female. These subjects reported no speech or hearing disorders. Their ages ranged from 18 to 35, with the average age being 22.3. They were recruited from the community at and around UC Berkeley and thus all spoke English as a second language. They all spent the majority of their childhood in South Korea, and all but one spoke the Seoul dialect of Korean (one subject's reported hometown was Ulsan, where the Gyeongsang dialect is found). The speaker who produced the tokens used for making stimuli did not participate in the experiment. All subjects were paid for their participation. Of the 14 subjects who participated in the experiment, one subject's data were excluded; this subject responded 'B' 158 times in a row early in the experiment.

D. Statistical analysis

1. Logit models and sensitivity

Data were analyzed using a logit mixed effects model fit with the lme4 package for R (Bates 2007). Logit models express how the likelihood of a binary response, in the form of a log odds ratio (*logit*), varies according to stimulus properties. Applying such a model to classification data involves using main effects to measure false alarms and interactions to measure the differences between hits and false alarms; the latter type of effect is a measure of *sensitivity*, similar but not identical to the d' measure of Signal Detection Theory (Macmillan & Creelman 1991).

In a model with ‘B response’ as the dependent variable, for instance, a main effect of stimulus type would estimate the difference between the logit of ‘B’ responses to stimulus-type B, i.e. hits, and the logit of ‘B’ responses to stimulus type A, i.e. false alarms. This difference in likelihood of hits and false alarms is a measure of sensitivity: the larger the difference, the more likely listeners are to label B stimuli as B relative to labeling A stimuli as B. If the stimuli are not discriminable, subjects will be equally likely to respond ‘B’ to either type and the parameter will be equal to 0; this is chance performance. If the stimuli are discriminable, hits will be more likely than false alarms, and the parameter will be greater than 0. To compare *differences* in sensitivity between different conditions, the model uses interactions between stimulus type and condition, estimating how the sensitivity effects discussed above differ between conditions.

2. Fixed effects

Experimental manipulations (which stimulus is being played and which stimuli are being compared in the block) are modeled as fixed effects. They are reported here with the effect coefficient β , and a Z-statistic and p-value from the Wald test. Fixed effects for experimental condition were dummy-coded for pairwise comparisons along the scale *simple F0 contrasts* < *simple voicing contrasts* < *oppositional contrast* < *integrative contrast*. The most important prediction of the cue-integration hypothesis is that the integrative contrast should be easier to perceive than the oppositional one, and subjects should therefore display greater sensitivity in the integrative-contrast condition. The simple-contrast conditions were included as a kind of control, to ensure that subjects could do the task and to compare with their performance in the oppositional and integrative conditions.

The dependent variable was ‘subject responded B’. For each condition, sensitivity parameters were estimated by using stimulus type (A or B) as an independent variable. Thus, the interaction between condition and stimulus type measures sensitivity (‘B’ hits minus ‘B’ false alarms) in the given condition relative to that in a baseline condition. In the scalar coding used here, the baseline is the condition immediately below in the scale. For instance, the interaction of oppositional-contrast condition with stimulus type estimates the difference in sensitivity between the oppositional-contrast condition and the simple voicing conditions.

The model also included fixed effects for several task-related variables that seemed likely to impact performance. ‘Block’ indicates the ordinal block (out of 6 in the experiment) during which the stimulus occurred, to capture fatigue and/or acclimation effects; it was coded as an orthogonal polynomial. ‘Post-error’ indicates whether the trial in question followed an incorrect answer on the previous trial. ‘Post-switch’ indicates whether the stimulus in the trial in question was different from the stimulus in the previous trial. These task effects were checked for interactions with stimulus type to examine how they affected sensitivity rather than just bias.

3. Random effects

Mixed models allow us to generalize across levels of random variables, variables sampled from a larger population which are not themselves the primary object of investigation (Jaeger 2008). In this study, we generalize across subjects by including random intercepts and by-subject random slopes for effects of interest. The general idea is that the model assumes subjects may vary with regard to patterns of response bias and sensitivity, and assesses the reliability of effects taking into account this variation. By-subject random slopes were added to the model in stepwise fashion, starting with the largest simple effects and progressing through all significant

interactions. Only random effects below the significance-level of $\alpha = 0.05$ were retained in the model. Significant by-subjects effects are reported here with the chi-square statistic, degrees of freedom, and p-value from a likelihood-ratio test. The chi-square statistic is a measure of how much the effect in question improves model fit.

III. RESULTS

A. Fixed effects

The sensitivity parameters fit by the model are shown in figure 3, in the scalar order in which condition variables were coded. Sensitivity to simple F0-onset contrasts, in which stimuli differed only in the F0 onset of the post-fricative vowel, were compared to chance (0); listeners performed significantly better than chance on these contrasts: $\beta = 1.25$, $Z = 3.35$, $p < 0.001$.

Listeners performed significantly better on simple voicing contrasts, where stimuli differed only in the presence vs. absence of a voice bar in the last 75% of the fricative, than the F0-onset contrasts: $\beta = 0.45$, $Z = 2.20$, $p = 0.028$. Listeners performed slightly worse on the oppositional contrast, where low F0-onset correlated with devoicing, than the simple voicing contrasts, but this trend is not significant. Finally, and most importantly for the experimental hypothesis, listeners performed significantly better on the integrative contrast, where low F0-onset correlated with voicing, than the oppositional one: $\beta = 2.09$, $Z = 2.65$, $p < 0.01$.

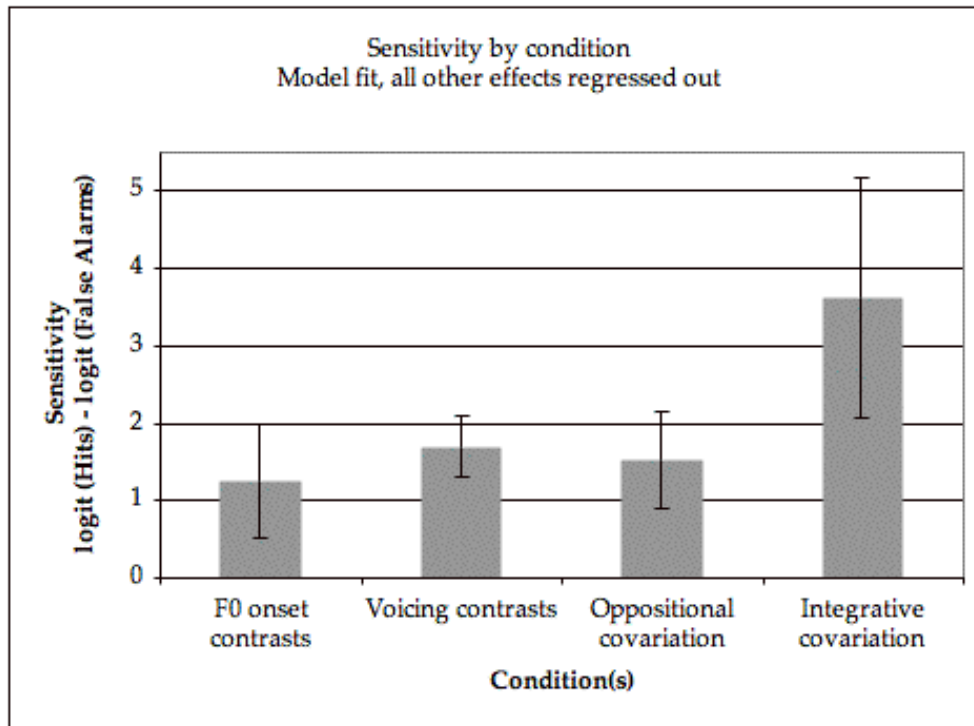


FIG. 3. Sensitivity by type of contrast, as determined by the logit mixed model. Error bars show 95% confidence intervals. Chance performance is 0.

Two of the three task-related variables included in the model significantly impacted sensitivity. Linear and quadratic terms for block were both significant. For the linear term, $\beta = 0.93$, $Z = 2.65$, $p = 0.050$; for the quadratic term, $\beta = -1.10$, $Z = 2.93$, $p < 0.01$. Inspection of the data suggests that most subjects increased in sensitivity throughout the first half of the experiment, ostensibly a training effect; then decreased in sensitivity throughout the second half, ostensibly a fatigue effect. The two significant terms, one positive and one negative, reflect this broad profile. Post-switch was also significant: subjects performed worse on trials where the stimulus was different from the preceding trial: $\beta = -0.47$, $Z = 3.54$, $p < 0.001$. Post-error was not significant.

B. Random effects

Several parameters showed significant variation across subjects. Overall sensitivity varied by subject, and incorporating this variation significantly improved model fit: $\chi^2 = 648$ on 2 d.f., $p < 0.001$. Subjects also differed with regard to the effect of block on sensitivity, and parameters for this difference significantly improved model fit: $\chi^2 = 300$ on 52 d.f., $p < 0.001$. Finally, subjects differed in the magnitude (but not the existence) of the sensitivity advantage for integrative contrast over other contrasts: $\chi^2 = 36$ on 23 d.f., $p = 0.044$. All of the results for fixed effects reported above come from a model that takes these by-subject differences into account; we can therefore conclude that the significant fixed effects are robust to between-subjects variation.

IV. DISCUSSION

The results of the experiment suggest that low F0 onset in a following vowel and the presence of voicing during frication integrate perceptually. By hypothesis, the two cues both contribute to the intermediate perceptual property of increased low-frequency energy. When both cues contribute to this property, stimuli are easier to discriminate. These results replicate for Korean listeners and fricatives Kingston *et al.*'s (2008) results for English listeners and stops.

In addition to providing evidence that the two cues integrate, the current study also provides evidence that prior experience with acoustic covariance is not *necessary* for integration to occur (Kingston *et al.* (2008) demonstrated that such experience is not *sufficient*). Given that there is no evidence that the two fricatives in Seoul Korean differ in their effects on the F0 onset of a following vowel, Seoul Korean speakers are unlikely to have experience with voicing and F0 covarying in this way in fricatives. It is thus unlikely that this particular kind of cue integration is learned from the linguistic input.

One possible objection to this argument is that listeners may be generalizing their knowledge of the laryngeal properties of *stops* to fricatives. Lenis stops are associated with significantly lower F0 onset in a following vowel than aspirated and fortis stops, and the lenis series is realized as voiced in between vowels (Cho, Jun, & Ladefoged 2002). Perhaps listeners have learned from medial lenis stops that voicing and low F0 tend to covary, and have generalized this knowledge to fricatives. One serious difficulty exists for this interpretation, however: Chang (2013) finds that the perceptual effect of low F0 onset exists for *initial* non-fortis fricatives, which are strongly aspirated. Because aspirated stops are realized with slightly higher F0 onset in a following vowel than fortis stops, listeners would have to be suppressing generalizations about F0 from a phonetically similar category (aspirated stops) in favor of generalizations from a phonetically dissimilar category (lenis stops). It is unclear what could drive such a mechanism.

One more result deserves mention: we do not find statistically significant evidence that sensitivity decreases when the two cues contribute to the intermediate percept in opposite directions (oppositional contrast) relative to varying only one cue. There was, however, a non-significant trend in this direction. As with most negative findings, it is hard to draw any firm conclusions from this. It may indicate that the relevant notion of ‘integration’ is asymmetric in an interesting way, with no interference when components of an integrative property differ in oppositional ways. But the results are also consistent with the existence of an interference effect that is too small to be reliably detected in our experiment.

This study has answered one small question about the perception of the laryngeal contrast for Korean fricatives; many questions still remain. In particular, it would be interesting to test whether other cues implicated in the laryngeal contrast for fricatives also integrate with the two

investigated here. Both Cho, Jun & Ladefoged (2002) and Chang (2013), for instance, report that the non-fortis fricative displays higher amplitude of the first harmonic in the following vowel relative to the second harmonic, an acoustic feature associated with breathy voice. As this property will tend to increase the amount of low-frequency energy in the signal, the theory advanced here predicts that it should integrate with voicing and low F0 onset. More generally, hypotheses about cue-integration may help make sense of the unusual laryngeal contrasts in Korean, for stops as well as fricatives: the three-way stop contrast involves a complex mix of duration, voice quality, VOT, and burst cues that differ between prosodic positions (Cho & Keating 2001, Cho, Jun & Ladefoged 2002). Examining this heterogeneous set of acoustic properties in terms of higher-level intermediate perceptual properties may offer a more unified way to think about the various cues involved.

In terms of the two models of cue integration discussed earlier, this study provides support for the existence of *auditory integration* above and beyond (or instead of) *empiricist integration*. This is because the cues that were shown to integrate perceptually for Korean listeners are not cues that generally covary in Korean production. Coupled with Kingston *et al.*'s (2008) results showing that production covariance is not sufficient for perceptual integration, this suggests that the integration of acoustic cues must be less constrained in some ways than the empiricist approach predicts (because not every instance of production covariance results in integration) and more constrained in other ways (because there are instances of integration that do not correspond to production covariance). Intermediate perceptual properties and auditory integration provide a promising starting point for thinking about what the relevant constraints may be.

REFERENCES

- Abramson, A. and Lisker, L. (1973). Voice timing in Korean stops. In *Proceedings of the seventh international congress of phonetic sciences, Montreal*, 439-446.
- Chang, C. (2013). The production and perception of coronal fricatives in Seoul Korean. *Korean Linguistics* **15**(1), 7-49.
- Cho, T., Jun, S., and Ladefoged, P. (2002). Acoustic and aerodynamic correlates of Korean stops and fricatives. *J. Phon.* **30**(2), 193-228.
- Cho, T., and Keating, P. (2001). Articulatory and acoustic studies on domain-initial strengthening in Korean. *J. Phon.* **29**, 155-190.
- Dart, S. (1987). An aerodynamic study of Korean stop consonants: Measurements and modeling. *J. Acoust. Soc. Am.* **81**(1), 138-147.
- Flanagan, J.L. & Saslow, M.G. (1958). Pitch discrimination for synthetic vowels. *J. Acoust. Soc. Am.* **30**, 435-442.
- Garner, W. R. (1974). *The processing of information and structure*, 128-131. Potomac, MD: Lawrence Erlbaum Associates.

Han, M., and Weitzman, R. (1970). Acoustic features of Korean /P, T, K/, /p, t, k/, and /ph, th, kh/. *Phonetica* **22**, 112-128.

Jaeger, T.F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *J. Mem. Lang.* **59**(4), 434- 446.

Kim, H., Maeda, S., Honda, K., and Hans, S. (2010). The Laryngeal Characterization of Korean Fricatives: Acoustic and Aerodynamic Data. In Fuchs, Toda, & Zygis (eds.), *Turbulent Sounds: An Interdisciplinary Guide*, 143–166. Berlin: Mouton de Gruyter.

Kingston, J., & Diehl, R. (1994). Phonetic knowledge. *Language* **70**, 419–454.

Kingston, J., Diehl, R., Kirk, C., and Castleman, W. (2008). On the internal perceptual structure of distinctive features: The [voice] contrast. *J. Phon.* **36**(1), 28-54.

Klatt, D.H. (1973). Discrimination of fundamental frequency contours in synthetic speech: Implications for models of speech perception. *J. Acoust. Soc. Am.* **53**, 8-16.

Kluender, K. (1994). Speech perception as a tractable problem in cognitive science. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics*, 173–217. San Diego: Academic Press.

Kurowski, K., and Blumstein, S. (1984). Perceptual integration of the murmur and formant transitions for place of articulation in nasal consonants. *J. Acoust. Soc. Am.* **76**(2), 383-390.

Liberman, A., Delattre, P., Cooper, F., and Gerstman, L. (1954). The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychol. Mono.* **68**, 1-13.

Liu, C. (2013). Just noticeable difference of tone pitch contour change for English- and Chinese-native listeners. *J. Acoust. Soc. Am.* **134**(4), 3011-3020.

Macmillan, N.A., and Creelman, C.D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.

Malecot, A. (1956). Acoustic cues for nasal consonants: an experimental study involving a tape-splicing technique. *Language* **32**(2), 274-284.

Nearey, T. M. (1997). Speech perception as pattern recognition. *J. Acoust. Soc. Am.* **101**, 3241–3254.

Parker, E., Diehl, R., and Kluender, K. (1986). Trading relations in speech and nonspeech. *Perception and Psychophysics* **39**, 129–142.

Silverman, D. & Jun, J. (1994) Aerodynamic evidence for articulatory overlap in Korean, *Phonetica* **51**, 210-220.

t'Hart, J. (1981). Differential sensitivity to pitch distance, particularly in speech. *J. Acoust. Soc. Am.* **69**, 811-821.

Yoon, K. (1999). A study of Korean alveolar fricatives: An acoustic analysis, synthesis, and perception experiment. MA thesis, University of Kansas.

FIGURE CAPTIONS

FIG 1. Schematic illustration of two acoustic dimensions, x and y , that integrate during the course of perception. The contrast between the stimulus where both values are high and the one where both values are low is *integrative*: the cues work together and the contrast is more perceptible. The contrast between the stimuli where one value is high and one low is *oppositional*: the cues work against each other and the contrast is less perceptible.

FIG. 2. Spectrograms and (smoothed) F0 contours of the four stimuli used in the experiment. The two in the top row are voiced, the two in the bottom row devoiced. The two on the left have higher F0 onset at the end of frication, the two on the right lower F0 onset.

FIG. 3. Sensitivity by type of contrast, as determined by the logit mixed model. Error bars show 95% confidence intervals.