Dynamic aspects of the production and perception of Korean sibilant fricatives

Jeffrey J. Holliday[1], Patrick F. Reidy[2]
[1] Korea University; [2] Waisman Center; Department of Communication Sciences & Disorders, University of Wisconsin-Madison

Introduction. Studies of sibilant fricatives have traditionally treated their spectra as static things, rather than as evolving in time across the duration of frication. For example, spectra of English /s/ and /ʃ/ are typically estimated from a single interval and then treated as a discrete probability mass function from which its moments (e.g., centroid, skewness) are computed (e.g., Li et al., 2011; Romeo et al., 2013). Other studies have computed spectral moments from multiple locations throughout the fricative, but did not analyze how the moments varied from location to location (e.g., Jongman et al., 2000); hence, any temporal variation in spectral properties has gone unexamined. Recently, it was demonstrated that the centroid frequency of English /s/ varied substantially across its duration (Iskarous et al., 2011). Furthermore, dynamic spectral measures, may reveal acoustic distinctions where static measures show none: For example, English and Japanese /s/ are indistinguishable in terms of static peak frequency at fricative midpoint, but *are* distinguishable in terms of the linear and quadratic trends of their peak frequency trajectories (Reidy, 2015).

Acoustic analyses of the Korean sibilant fricative contrast, /s^h/-/s*/, have also relied almost entirely on static spectral measures (e.g. Chang, 2013; Kallay & Holliday, 2012), despite the fact that the acoustic profile of Korean /s^h/ changes substantially over the course of the fricative. Specifically, when followed by a non-high vowel, /s^h/ is typically aspirated, resulting in an earlier release of the lingual closure than in /s*/. This difference in articulation has previously been indexed in terms of acoustic features (e.g. F1 or H1-H2) of the vowel onset (e.g. Chang, 2013). The aims of the present study are to explore how the Korean /s^h/-/s*/ contrast can be quantified using dynamic acoustic measures computed within the fricative, and to show how listeners' identification and goodness judgements of fricative productions correlate with such measures.

Method. 6 native Korean speakers and 6 native Mandarin L2 learners of Korean (all female) produced 18 fricative-initial words, which included 3 words for each CV combination of /s^h, s*/ × /a, u, i/. Each CV was extracted from the word production, RMS normalized to 65 dB, and then measurements were made in the fricative and at vowel onset. Centroid frequency was measured from 17 multitaper spectra estimated from 20-ms windows spaced evenly across each token, and a quadratic orthogonal polynomial model was fitted to the 17 centroid values. The fitted coefficient of the model's linear term ("linear centroid") and centroid frequency at midpoint ("midpoint centroid") were used to characterize the fricative. H1-H2 was measured from an FFT of a 25-ms window left-aligned at fricative-vowel boundary. Only the measurements made on the tokens produced by the native speakers were used in the production analysis.

We then conducted a perception experiment in which 12 native Korean listeners identified the category and rated the goodness of the fricative productions of both the native Korean and native Mandarin speakers. The productions by L1 Mandarin speakers were included to ensure that a range of goodness ratings would be obtained. In the identification portion of the experiment stimuli were blocked by vowel, and in the goodness rating portion stimuli were blocked by CV (e.g. listeners rated all of the /s^{h}a/ tokens, then all of the /s^{h}u/ tokens, etc). Then, the entire procedure was repeated with stimuli consisting of only the initial C, with the following vowel removed.

Results. Following Chang (2013), a repeated measures ANOVA was built for each acoustic measure, with the measure as the DV and fricative target and vowel as within-subject IVs. Following Holliday et al. (2015), we then calculated the percentage of tokens correctly predicted (%CP) by a logistic mixed-effects model, once with fricative category as the nesting factor, and once with both fricative category and vowel context as nesting factors. The ANOVA showed that all measures were predictive of fricative category. As far as %CP, H1-H2 (77.8%) and midpoint centroid (76.9%) were the best predictors, although linear centroid performed just as well (75.9%) when vowel context was built into the model.

Identification accuracy rates for the native Korean targets were above chance ($p < 0.001$) in all vowel contexts, with the highest accuracy for /a/ in the CV condition (96.5%) and the lowest for /u/ in the C condition (63.7%). Accuracy for /a/ was high even in the C condition (85.4%), suggesting that the cues in the fricative alone may be sufficient for identification in this vowel context. Listeners were overall biased towards /s^h/, choosing it on 58.6% of the native Korean stimuli trials. Listeners were especially biased toward /s^h/ in the C condition, choosing it on 64.1% and 67.8% of the /a/ and /i/ trials, respectively.

Then, for each stimulus, and for both the CV and C conditions, we calculated the proportion of /s*/ responses in the identification task and the mean goodness rating in the rating task. We used the proportion of /s*/ responses instead of identification accuracy because the acoustic measures make opposite predictions about accuracy for/s^h/ and /s*/. We then regressed these identification and goodness measures against the acoustic measures (except for H1-H2 in the C condition). In the CV condition, /s*/-response in the /a/ context was highly correlated with H1-H2 ($R^2 = .72$), linear centroid ($R^2 = .61$), and midpoint centroid ($R^2 = .58$). Goodness ratings of both /s^{h}a/ and /s*a/ were similarly correlated with all three of these measures ($.49 < R^2 < .59$). In the /u/ and /i/ contexts, where the fricative spectrum is less dynamic, linear centroid was less predictive of either /s*/-response or goodness. In the C condition, linear centroid was more predictive than midpoint centroid of /s*/-response in the /a/ context ($R^2 = .69$ vs. $R^2 = .53$), but far less so in the /i/ context ($R^2 = .18$ vs. $R^2 = .40$). Neither measure was highly predictive of /s*/-response in the /u/ context ($R^2 < .13$). Goodness ratings in the C condition displayed a similar pattern as in the CV condition.

In summary, it was found that the linear coefficient (essentially, the slope) of the trajectory of centroid frequencies across the frication of /s^h/ was at least as good a predictor of category response in the /a/ context as the static measure of centroid, but not necessarily a better predictor than H1-H2 when the full CV was provided. This trend held in the C condition as well, when H1-H2 was unavailable. This result suggests that when Korean listeners categorize /s^{h}a/ and /s*a/ they may weigh dynamic factors more than static ones. That is, what matters for /s^{h}a/-/s*a/ perception is not simply the frequency range within which the frication noise is concentrated, but the rate at which the frequency range drops as the lingual constriction is released.

References. Chang, C. B. (2013). "The production and perception of coronal fricatives in Seoul Korean." *Korean Linguistics*, 15, 7-49. // Holliday, J. J., Reidy, P. F., Beckman, M. E., & Edwards, J. (2015). Quantifying the robustness of the English sibilant fricative contrast in children." *J. Speech Lang. Hear. Res.*, 58, 622-637. // Iskarous, K., Shadle, C. H., & Proctor, M. I. (2011). "Articulatory-acoustic kinematics: The production of American English /s/." *J. Acoust. Soc. Am*., 129(2), 944-954. // Jongman, A., Wayland, R., & Wong, S. (2000). "Acoustic characteristics of English fricatives." *J. Acoust. Soc. Am.*, 108(3), 1252-1263. // Kallay, J. & Holliday, J. J. (2012). "Using spectral measures to differentiate Mandarin and Korean sibilant fricatives." *Proceedings of INTERSPEECH 2012*. // Li, F., Edwards, J., & Beckman, M. E. (2011). "Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers." *J. Phon.*, 37, 111-124. // Reidy, P. F. (2015). The spectral dynamics of voiceless sibilant fricatives in English and Japanese. Unpublished

Ph.D. dissertation, The Ohio State University. // Romeo, R., Hazan, V., and Pettinato, M. (2013). "Developmental and gender-related trends of intra-talker variability in consonant production." *J. Acoust. Soc. Am.*, 134(5), 3781-3792. // Thomson, D. J. (1982). "Spectrum estimation and harmonic analysis." *Proceedings of the IEEE*, 70, pp. 1055-1096.