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; Dept. of Communication Sciences & Disorders, Univ. of Wisconsin-Madison

Background: Sibilant fricatives

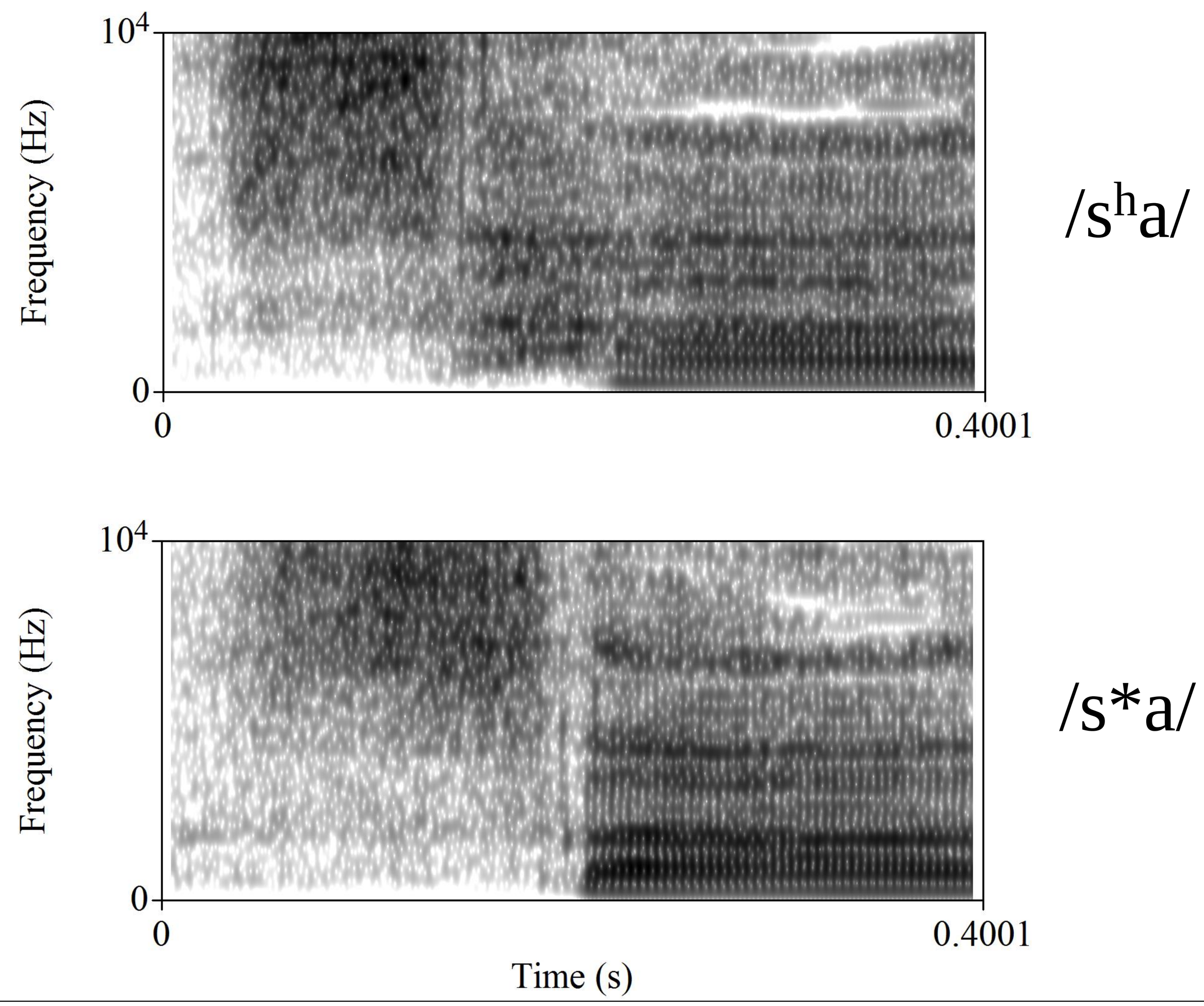
- Most studies of sibilant fricatives have treated their spectra as static.
  - Spectral moments at midpoint (Li et al., 2009; Romeo et al., 2013)
  - Spectral moments at several time points, but no analysis of how the moments varied over time (Jongman et al., 2000)
- But sibilant fricative spectra do change over the course of the fricative (e.g. Iskarous et al., 2011).
  - English and Japanese /s/ do not differ in peak ERB frequency, but do differ in terms of peak frequency trajectory across the fricative (Reidy, 2015).

Background: Korean fricatives

- Korean has two sibilant fricatives, /sʰ/ and /s*/, which differ greatly in spectral dynamics, but have analyzed almost exclusively using static measures (e.g. Chang, 2013; Kallay & Holliday, 2012).
  - /sʰ/ has an earlier release of the lingual closure, resulting in aspiration before low and mid vowels (/a/, /ɛ/, /o/, /ʌ/), but not before high vowels (/i/, /ɨ/, /u/).
  - /s*/, however, is never aspirated.
  - /sʰ/, and sometimes /s*/, is palatalized before /i/ (e.g. /sʰi/ → [ɕi])
- Commonly used acoustic measures include F1 and H1-H2 at the onset of the following vowel, and centroid frequency at some point during the frication.

Research questions

1. Could Korean /sʰ/ and /s*/ be better differentiated using dynamic rather than static acoustic measures?
2. Would listeners' perception of these fricatives be better predicted by differences in dynamic rather than static measures?



Method & Analysis: Production

Participants

- 6 female native Korean speakers

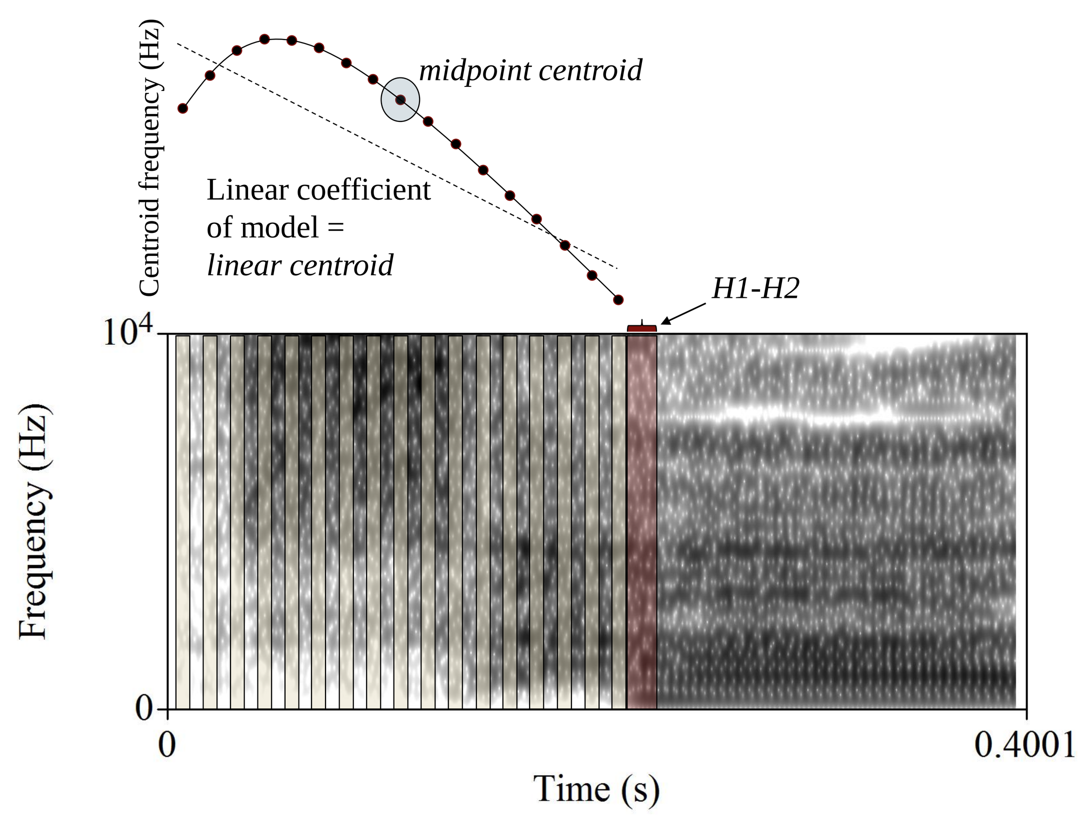
Stimuli

- 18 fricative-initial words = {sʰ, s*} + {a, i, u} × 3 words
- Word-initial CVs extracted, RMS normalized to 65 dB

Measurements

- Centroid measured from 17 multitaper spectra estimated from 20-ms windows evenly spaced across the fricative.
- Quadratic orthogonal polynomial model fit to the 17 centroid values
- Three measures used in analysis:
  1. Coefficient of model's linear term ("linear centroid")
  2. Centroid from fricative midpoint ("midpoint centroid")
  3. H1-H2 from a 25-ms window taken at vowel onset

Illustration of acoustic measures



Results: Production

RM ANOVA (measure ~ fricative*vowel)

All acoustic measures revealed a main effect of fricative category:

- Linear centroid: $F(1,5) = 25.4$, $p = .004$
- Midpoint centroid: $F(1,5) = 51.6$, $p < .001$
- H1-H2: $F(1,5) = 13.6$, $p = .0142$

Fricative-vowel interaction terms were significant as well:

- Linear centroid: $F(2,10) = 17.3$, $p < .001$
- Midpoint centroid: $F(2,10) = 18.3$, $p < .001$
- H1-H2: $F(2,10) = 5.8$, $p = .021$

Method & Analysis: Perception

Participants

- 12 native Korean listeners

Stimuli

- Extracted from the word productions of the 6 native Korean speakers above, plus productions from 6 female native Mandarin L2 learners of Korean (L2 productions included to ensure a wide range of goodness ratings).
- Full CV, and C only (with V removed)

Procedure

- Full CV stimuli
  - Identify the fricative category – blocked by V
  - Provide a goodness rating – blocked by CV
- C only stimuli
  - Same procedure as above, but the vowel portion was removed from the stimuli. Listeners were told what the following vowel originally was, however.

Results: Perception

Identification accuracy

	Vowel context	Accuracy	/sʰ/ response rate
Full CV	/a/	96.5%	53.0%
	/i/	83.6%	55.8%
	/u/	83.1%	52.6%
C only	/a/	85.4%	64.1%
	/i/	65.0%	67.8%
	/u/	63.7%	58.1%

- As in previous studies, identification accuracy is much poorer in high vowel contexts.
- While vocalic cues seem to carry most of the information needed for the /i/ and /u/ contexts, listeners can identify /sʰa/-/s*a/ reasonably well without any vocalic information at all.
- Listeners are biased toward /sʰ/ when no vocalic cues are present.

Relationship between identification, goodness ratings, and acoustic measures

R^2 of linear models predicting identification and goodness ratings

Identification	**Full CV**			**C only**		
	/a/	**/i/**	**/u/**	**/a/**	**/i/**	**/u/**
Linear centroid	**.613**	.106	.058	**.690**	.181	.128
Midpoint centroid	**.578**	**.371**	*n.s.*	**.527**	**.398**	.106
H1-H2	**.718**	*n.s.*	*n.s.*			
Goodness rating	**Full CV**			**C only**		
	/sʰa/	**/sʰi/**	**/sʰu/**	**/sʰa/**	**/sʰi/**	**/sʰu/**
Linear centroid	**.586**	**.419**	.097	**.566**	.288	.173
Midpoint centroid	**.508**	.262	*n.s.*	**.503**	**.463**	.092
H1-H2	**.541**	.146	*n.s.*			
	/s*a/	**/s*i/**	**/s*u/**	**/s*a/**	**/s*i/**	**/s*u/**
Linear centroid	**.494**	*n.s.*	*n.s.*	**.648**	*n.s.*	*n.s.*
Midpoint centroid	**.498**	.106	.208	**.370**	**.529**	.173
H1-H2	**.548**	.168	*n.s.*			

Linear centroid predicts identification response very well in the /a/ context, but is less predictive than midpoint centroid in the /i/ context (when aspiration is reduced and /sʰ/ is palatalized to [ɕ]).

Linear centroid also predicts perceived goodness very well in the /a/ context, but less so in high vowel contexts.

Conclusion: Vocalic cues (e.g. H1-H2) do seem to be primary, but explaining identification accuracy across vowel contexts when vocalic cues are absent may be best explained by a combination of both dynamic and static cues.

References

Chang, C. B. (2013). The production and perception of coronal fricatives in Seoul Korean. *Korean Linguistics, 15*, 7-49. // 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.