Vowel-context effects on the spectral dynamics of English & Japanese sibilant fricatives

Patrick F. Reidy & Mary E. Beckman

Department of Linguistics, Ohio State University

Purpose

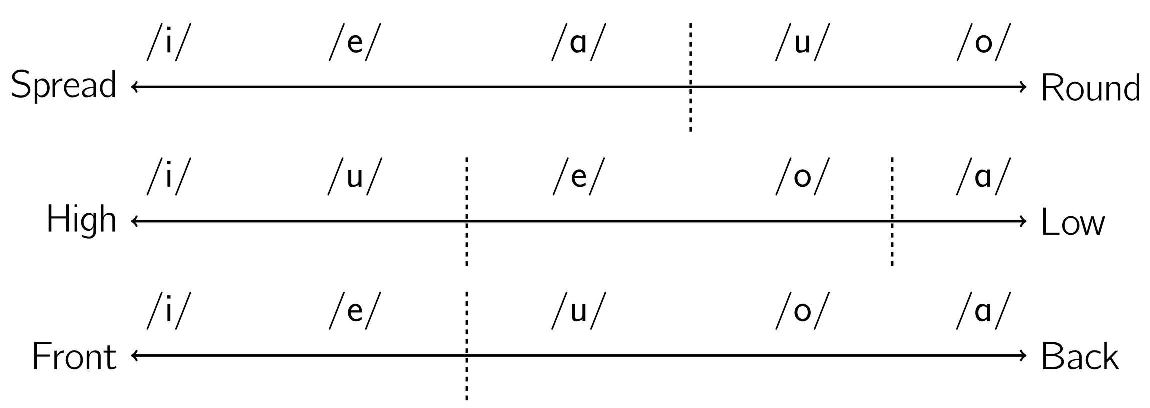
- ▷ Fluent speech requires the coproduction of neighboring speech sounds; thus, the acoustic properties of a speech sound are adapted to the context in which it is produced.
- ▷ Articulographic (Katz & Bharadwaj, 2001) and ultrasound (Zharkova et al., 2014) studies have found temporal coarticulatory effects of a following vowel on the trajectory of the tongue during the production of a sibilant.
- ▷ Anticipatory vowel-context effects on the spectral properties have been reported at discrete locations in the sibilant—e.g., near frication midpoint or vowel boundary (McGowan & Nittrouer, 1988).
- ▶ Spectral properties of sibilants vary temporally (Iskarous et al., 2011; Reidy, 2015); hence, the current study investigated effects of vowel context on the *spectral dynamics* of English /s/ and /ʃ/ and Japanese /s/ and /ɕ/.
- ▶ Two psychoacoustic spectral properties were considered: (1) *peak ERB_N-number*, a measure of the most prominent frequency; (2) *excitation drop*, a measure of spectral balance.

Background

Vowel features

- ▷ In English & Japanese, the vowels /i/, /u/, /e/, /o/, /ɑ/ exhibit language-specific placement along spread-round, high-low, and front-back continua (e.g., Japanese /u/ is relatively more spread than English /u/), but broadly follow the pattern shown in Fig. 1 (cf. Fromkin, 1964, Fukuda & Hiki, 1982; Wada et al., 1969).

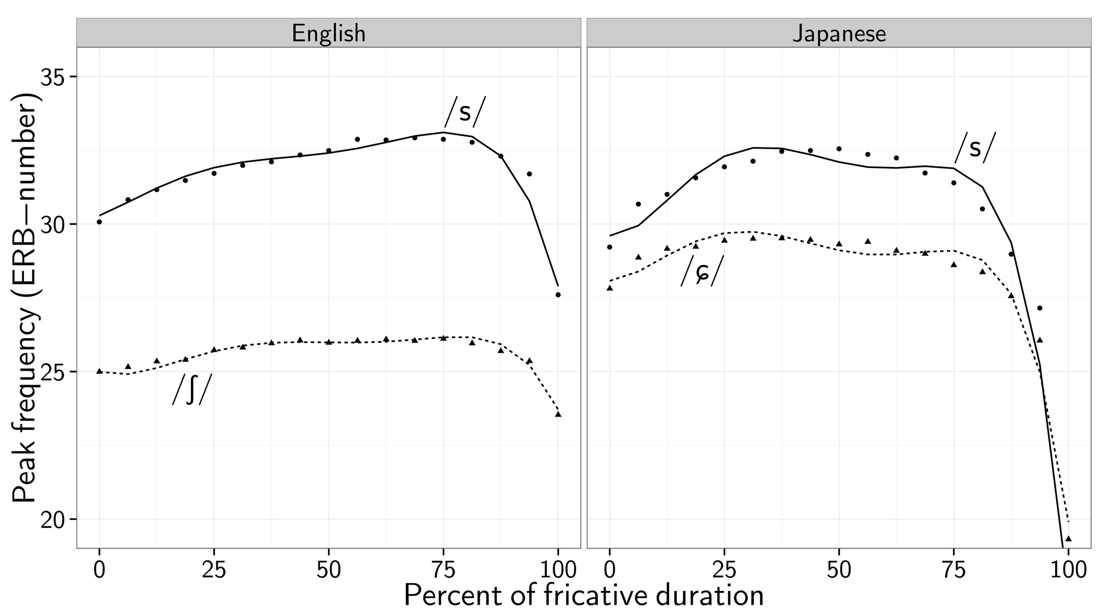
Figure 1: General placement of vowels along feature continua. Discretization of each continuum is marked by dotted lines.



Sibilant contrast

- ▷ English /s/ & /ʃ/: contrastive in all vowel contexts; /ʃ/ constriction more posterior, front cavity larger than /s/; /ʃ/ articulated with lip rounding.
- ▷ Japanese /s/ & /ɕ/: contrastive only in back vowel contexts; /ipaC/ constriction longer, more palatalized than /s/; both sibilants articulated with spread lips.
- ▷ In both languages, the sibilant contrast is indicated by static and dynamic aspects of spectral properties (cf. average trajectory level vs. trajectory shape in Fig. 2).
- ▷ Spectral resonances of English /s/ reported to be lower in round-vowel contexts, due to anticipatory lip rounding.

Figure 2: Temporal variation in peak ERB_N-number across the English and Japanese sibilants. Fixed-effects predictions of fifth-order polynomial growth curve models are shown as lines; data means, as points.



Methods

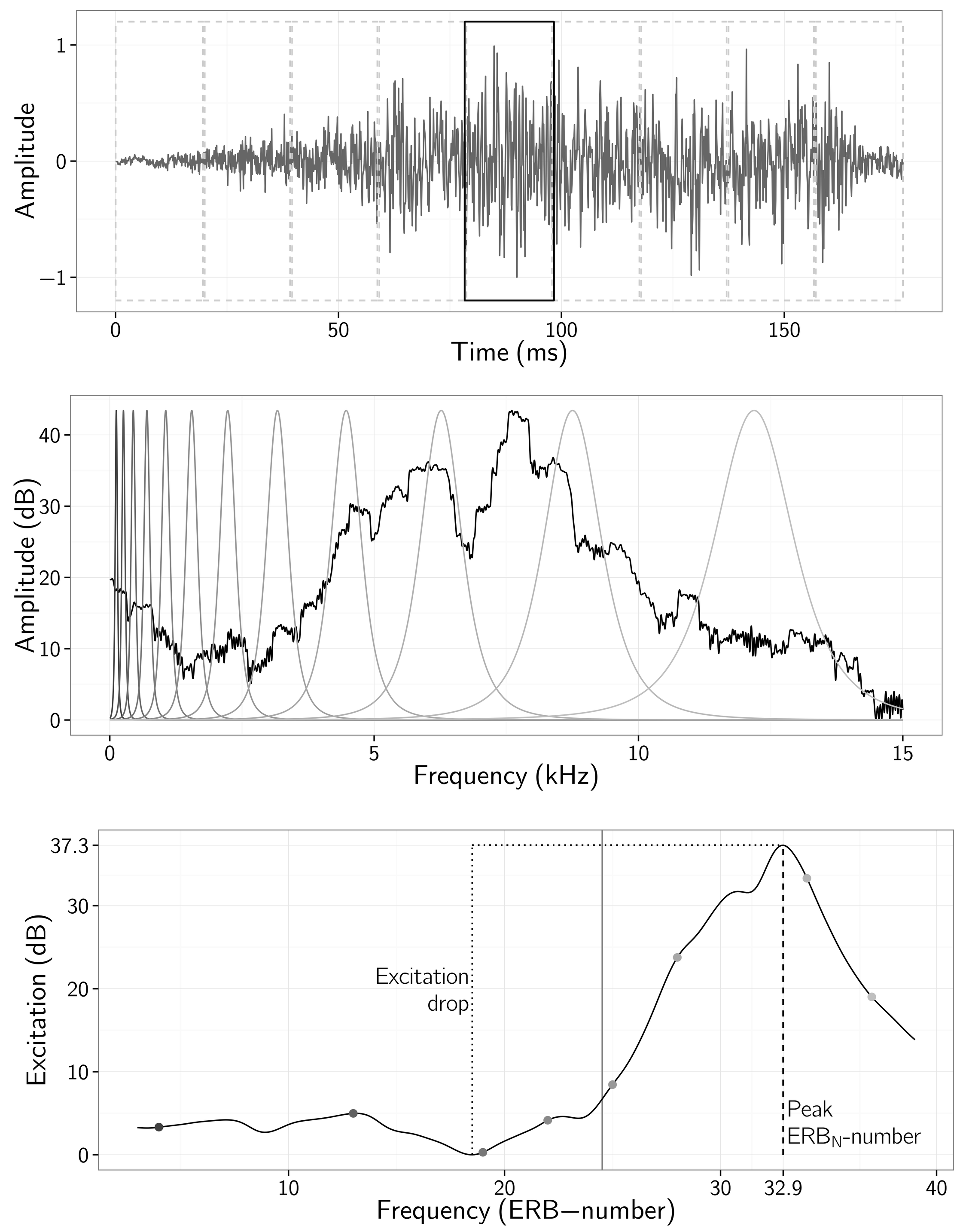
Participants & materials

- ▷ 20 adult native speakers (10 females, 10 males) for each language.
- ▷ Sibilants produced in prevocalic, initial position of real words.
  - – English vowels were grouped into five classes: {i}: /i, ɪ/; {u}: /u, ʊ/; {e}: /e, ɛ/; {o}: /o/; {ɑ}: /ʌ, ɑ, ɔ/.
  - – In each language, 3 target words per phonotactically legal combination of sibilant and vowel (class).
- ▷ Elicited with a picture-prompted repetition task.

Estimation of psychoacoustic spectral properties

- ▷ From each sibilant production, 17 excitation patterns were computed from 20-ms intervals spaced evenly across the frication (cf. Fig. 3).
- ▷ Psychoacoustic spectral properties computed from each excitation pattern:
  - ▶ *Peak ERB_N-number:* most prominent psychoacoustic frequency;
  - ▶ *Excitation drop:* difference in excitation (dB) between high-freq. peak and low-freq. trough (cutoff = 24.5 ERB_N-num. ≈ 3 kHz).
- ▷ Hence, each production represented with two 17-point trajectories.

Figure 3: Top: Waveform of /s/; odd-numbered 20-ms analysis windows overlaid. Middle: Multitaper spectrum estimated from middle window of /s/; gammatone filters overlaid. Bottom: Excitation pattern output by a 361-channel gammatone filter bank model of auditory periphery.



Analysis of vowel-context effects on spectral dynamics

- ▷ For each sibilant and spectral property, an orthogonal polynomial growth curve model was fitted to the trajectories: 8 models in total.
  - – Fixed effects: Vowel, $Time^n$ up to n = 5, and $Time^n$ × Vowel interactions. Built up with a stepwise forward-selection protocol.
  - – Random effects: Intercept and each power of $Time^n$ for talker and for vowel-within-talker.
- ▶ Interactions between Vowel and non-zero powers of $Time^n$ indicate differences in spectral dynamics conditioned by vowel context. These are of primary interest.

Results

Figure 4: Fixed-effects predictions of the growth curve models fitted to the peak ERB_N-number trajectories. The effects of Vowel included in a model's structure are shown in the bottom left corner of each panel. Each model also included a fixed effect of intercept and of each power of $Time^n$ up to n = 5.

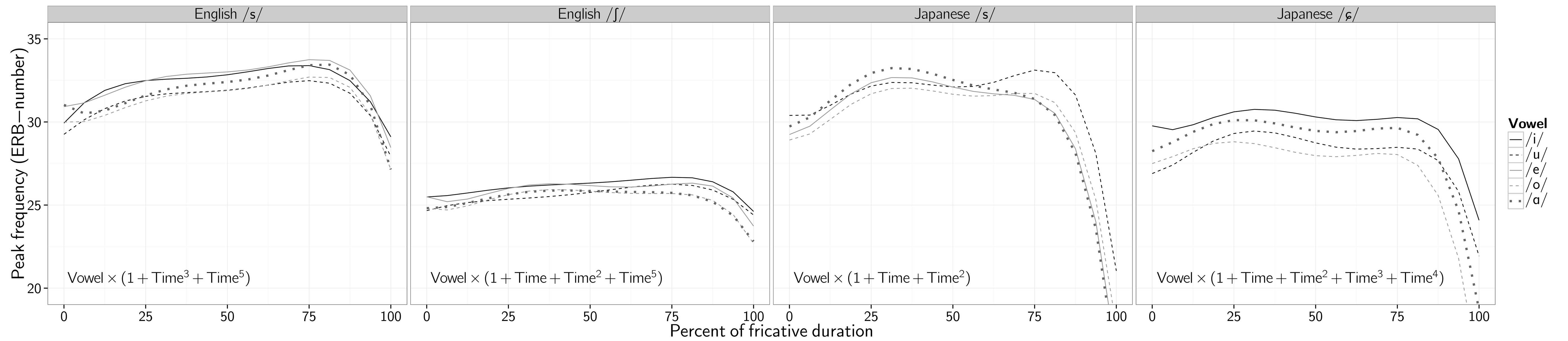
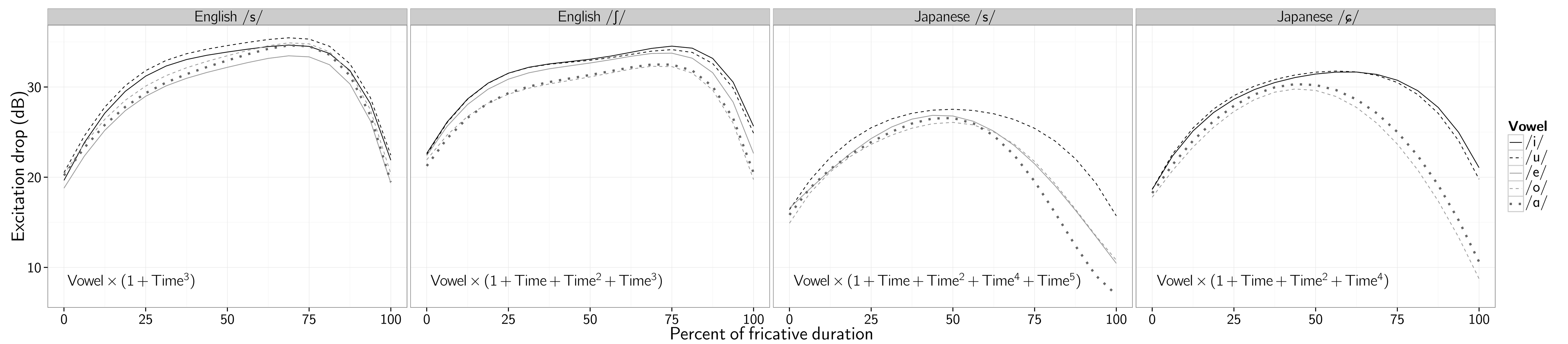


Table 1: Summary of Vowel coefficients in fitted models. Models were fitted with /ɑ/ as the reference level. Positive and negative effects are indicated by + and −, respectively. Significance was determined with Wald confidence intervals; shaded cells denote insignificant coefficients at the .05 level.

	Peak ERB_N-number													
	English								Japanese					
	/s/				/ʃ/				/s/			/ɕ/		
	/i/	/u/	/e/	/o/	/i/	/u/	/e/	/o/	/u/	/e/	/o/	/i/	/u/	/o/
Vowel	+	−	+	−	+	+	+	−	+	−	−	+	−	−
Vowel × Time					+	+	+	+	+	+	+	+	+	−
Vowel × $Time^2$					+	+	+	−	+	+	+	+	+	−
Vowel × $Time^3$	+	+	+	+								+	+	−
Vowel × $Time^4$												+	+	−
Vowel × $Time^5$	+	+	+	+	+	+	−	−						

	Excitation drop													
	English								Japanese					
	/s/				/ʃ/				/s/			/ɕ/		
	/i/	/u/	/e/	/o/	/i/	/u/	/e/	/o/	/u/	/e/	/o/	/i/	/u/	/o/
Vowel	+	+	−	+	+	+	+	−	+	+	+	+	+	−
Vowel × Time					+	+	+	−	+	+	+	+	+	−
Vowel × $Time^2$					+	+	+	+	+	+	+	+	+	−
Vowel × $Time^3$	+	+	+	+	+	+	+	−						
Vowel × $Time^4$									−	−	−	−	−	+
Vowel × $Time^5$									−	−	−			

Figure 5: Fixed-effects predictions of the growth curve models fitted to the excitation-drop trajectories. The effects of Vowel included in a model's structure are shown in the bottom left corner of each panel. Each model also included a fixed effect of intercept and of each power of $Time^n$ up to n = 5.



Discussion

- ▷ Significant main effects of Vowel indicated context effects on spectral statics (i.e., average level of spectral-property trajectory).
  - – Average peak ERB_N-number of English /s/ sensitive to vowel frontness and rounding; of English /ʃ/, only to vowel frontness.
  - – Average peak ERB_N-number of Japanese /ɕ/ sensitive to vowel frontness and rounding.
- ▷ Significant interactions between Vowel and non-zero powers of Time indicated context effects on spectral dynamics (i.e., shape of a spectral property's time course).
  - – Dynamics of peak ERB_N-number and exception-drop trajectories for both English /s/ and /ʃ/ were sensitive to vowel height.
  - – For Japanese /s/, dynamics of peak ERB_N-number and of excitation drop were sensitive to both vowel rounding (first half of frication) and vowel height (second half of frication). For Japanese /ʃ/, dynamics of peak ERB_N-number and of excitation drop were sensitive to height.
- ▶ The preponderance of context effects on non-zero powers of Time suggests the importance of jaw movement for the spectral dynamics of sibilant fricatives.

Acknowledgements

Data collection was supported by NIH Grant No. R01DC02932 to Jan Edwards (UW–Madison) and Mary E. Beckman. Data analysis was supported by a Stetson Scholarship from the ASA to Patrick F. Reidy.

References

- ▷ Fromkin, V. (1964). Lip positions in American English vowels. *Lang Speech*, 7(4):215–225.
- ▷ Fukuda, Y. & Hiki, S. (1982). Characteristics of the mouth shape in the production of Japanese: Stroboscopic observation. *J Acoust Soc Jpn*, 3(2):75–91.
- ▷ Iskarous, K. *et al.* (2011). Articulatory-acoustic kinematics: The production of American English /s/. *J Acoust Soc Am*, 129(2):944–954.
- ▷ Katz, W. F. & Bharadwaj, S. (2001). Coarticulation in fricative-vowel syllables produced by children and adults: A preliminary report. *Clin Linguist Phonet*, 15(1&2):139–143.
- ▷ Nittrouer, S. *et al.* (1989). The emergence of phonetic segments. *J Speech Hear Res*, 32:120–132.
- ▷ Reidy, P. F. (2015). The spectral dynamics of voiceless sibilant fricatives in English and Japanese. Ph.D. thesis, Ohio State University.
- ▷ Wada, T. *et al.* (1969). An approach for the cinefluorographic study of articulatory movements. *Proc 1969 Int Cong Cleft Palate*, 506–522.
- ▷ Zharkova, N. *et al.* (2014). Spatial and temporal lingual coarticulation and motor control in preadolescents. *J Speech Lang Hear R*, 57:374–388.