Steady as /ʃi/ goes: The spectral kinematics of sibilant fricatives in English and Japanese Patrick Reidy & Mary Beckman, The Ohio State University, Dept. of Linguistics

Motivation & Purpose of Current Study

- The voiceless sibilant contrasts of English $(/s/-/\int)$ and Japanese $(/s/-/\varsigma)$ have been extensively studied, and thus are well understood, in terms of these fricatives' *static* spectral features—e.g. peak frequency or spectral mean (centroid) at the temporal midpoint of frication.
- However, the production of a voiceless sibilant involves the continuous movement of the tongue and jaw—two articulators which participate in the generation of the fricative's noise sources, and which determine the geometry of the anterior cavity that is excited by those noise sources.
- Tongue: forms a narrow linguopalatal constriction; noise is generated as the airflow passing through the constriction becomes turbulent.
- Jaw: positions the incisors; noise is generated when turbulent airflow impinges on the incisors.
- The centroid of English /s/ follows an increasing, concave trajectory, which is partially accounted for by the rising-then-falling motion of the jaw during its articulation (Iskarous et al., 2011); however, no study has compared the spectral kinematics of multiple sibilants from more than one language.
- So, it is unknown whether the spectral kinematics of English /s/ are explained by general properties of the articulation of sibilant fricatives that hold either cross-linguistically or for English specifically.
- The current study investigates the spectral kinematics of English and Japanese voiceless sibilants as represented by the trajectory of a psychoacoustic spectral peak measure (*peak ERB*)—relative to the following research questions:
- 1. Do sibilants within a given language exhibit comparable peak ERB trajectories?
- 2. Do cross-linguistically assimilable sibilants display equivalent peak ERB trajectories?
- 3. Do English-acquiring children exhibit a developmental trajectory for differentiating /s and $/\int/$ in terms of their peak ERB trajectories?

Background

Adults' articulation of voiceless sibilant fricatives

• There is a binary voiceless sibilant contrast in each of English and Japanese. In both languages, /s/ is produced with a similar articulatory strategy. But, the articulatory strategy for the posterior fricatives $(/\int / and / \epsilon /)$ differs across the two languages.



- English lingual targets (top row)
- -/s/ (left): Tip flattened to make a dento-alveolar constriction. $-/\int/$ (right): Tip raised to make a post-alveolar constriction, resulting
- in a much larger front cavity than that of /s/. Japanese lingual targets (bottom row)
- -/s/ (left): Blade raised to make a dental constriction.
- $-/\varsigma/$ (right): Pre-dorsum bunched up to make a palatal constriction, resulting in a front cavity that is comparable in size, but with a different geometry from that of /s/.

(from Toda & Honda, 2003, p. 3)

Spectral features of sibilants at frication midpoint

- Due to differences in front cavity size or shape, in both English and Japanese, adults' productions of /s have higher resonances, relative to the posterior sibilant, resulting in higher centroid (e.g., Jongman et al., 2000; Li et al., 2009) and peak ERB (cf. left column)
- English-acquiring children initially produce /s, \int/as a single category, whose centroid H^{20} value at frication midpoint is closer to adults' /s than $/\int/$. Likewise, Japanese-acquiring children initially do not distinguish /s, $\wp/$; however, its centroid at frication midpoint is closer to what is appropriate for adults' $/\varsigma$ than /s (Li, 2012). This same crosslinguistic asymmetry is exhibited by the peak ERB of the children's productions (cf. right column).



Method

Participants, speech materials & elicitation procedure

- Native English- or Japanese-speaking adults, and three- to five-year-old children who were acquiring English natively participated in the study. (\approx 10 participants per language/age/gender.)
- Participants produced either English /s, $\int/$ or Japanese /s, g/ in word-initial, pre-vocalic position of real words in their respective native language, after being prompted by an audiovisual stimulus.
- Each production was transcribed, and only phonemically correct tokens were analyzed acoustically.

Computation of peak ERB trajectory

Waveform: English-speaking adult female's production of /s/ in 'soak'





Results: Adults' Peak ERB Trajectories





Results: Acquisition of Peak ERB Trajectories in English

Sibilant waveform pre-processing

- Onset and offset of frication marked manually by a trained phonetician.
- Waveform was not pre-emphasized.
- Nine 20-ms windows spaced evenly across duration of the sibilant.

Multitaper spectrum (K = 8; NW = 4)

- MTS is similar to DFT, but estimates ordinate values with less error.
- Spectra were estimated from the nine windows (light to dark). Spectral peak trajectory is shown as orange path.

Gammatone filterbank (361 channels)

- Center frequencies spaced every 0.1 ERB; bandwidths proportional to CFs.
- Models cochlear differential frequency selectivity, with respect to notchednoise masking.

Peak ERB trajectory

- The amplitude of the psychoacoustic spectrum at a given frequency ω is the total energy output by the filter, whose CF is ω , in response to a given input spectrum.
- Peak ERB trajectories (orange path) were normalized by subtracting the speaker's mean peak ERB of /s/.





Discussion

Sibilants of the same language do not necessarily exhibit similar peak ERB trajectories.

trajectory had greater curvature than that of $/\int/$ or $/\varsigma/$, respectively.

Cross-linguistically assimilable sibilants that have similar lingual target postures do not necessarily display comparable peak ERB trajectories.

- vowel may be timed differently for these two languages.

English-acquiring children must learn to differentiate /s/ and /ʃ/ in terms of their peak ERB trajectories.

seen for static spectral measures (cf. Li, 2012).

Acknowledgments

- Fellowship to the first author.
- Thanks are due to Fangfang Li for her contribution to this work.



• In both English and Japanese, adult speakers tended to produce /s/s such that its peak ERB

• The peak ERB trajectories of both English and Japanese /s/ were found to be concave, which would be expected given the rising-then-falling motion of the tongue and jaw that is necessary to form and then release the constriction during the articulation of a sibilant-vowel syllable.

• However, the peak ERB trajectory's concavity was much greater for Japanese than English /s/, which was largely due to the dramatic descent of the trajectory once it drops from its maximum.

• Furthermore, the peak ERB trajectory of /s/ reached its maximum later for English (window 6–7) than for Japanese (window 4–5), suggesting that the release of the constriction into the following

• The children produced $/\int/$ such that the shape of its peak ERB trajectory resembled that of the adults' /s/, although there was some evidence of the older female children flattening the peak ERB trajectory of $/\int/$, toward a more adult-like pattern—a developmental path that parallels that

• Work supported by NIDCD grant R01–02932 to Jan Edwards and by an OSU Graduate School