Learning curves: The acquisition of differential spectral kinematics of English sibilant fricatives

Patrick Reidy, *The Ohio State University, Dept. of Linguistics*

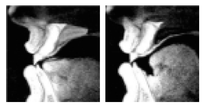
Motivation & Purpose of Current Study

- The voiceless sibilant contrast of English (/s/–/ʃ/) has been extensively studied, and thus is 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 either voiceless sibilant involves the *continuous movement* of the tongue and jaw—two articulators which participate in the generation of 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.
- If, due to their underlying articulatory dynamics, /s/ and /ʃ/ exhibit dissimilar spectral kinematic patterns, then it follows that static measurements of spectral features, like centroid or peak frequency, are incomplete measures of these features.
- The current study investigates the spectral kinematics of the English voiceless sibilants /s, ʃ/, as represented by the trajectory of a psychoacoustic spectral peak measure (*peak ERB*), relative to the following research questions:

1. **Do native English-speaking adults produce /s/ and /ʃ/ with comparable peak ERB trajectories?**
2. **Do English-acquiring children exhibit a developmental trajectory for distinguishing /s/ and /ʃ/ in terms of their peak ERB trajectories?**

Background

Adults' articulation of voiceless sibilant fricatives



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

Lingual targets

- /s/ (left): Tip flattened to make a dento-alveolar constriction.
- /ʃ/ (right): Tip raised to make a post-alveolar constriction, resulting in a much larger front cavity than that of /s/.

- During the articulation of /s/, the jaw rises and then falls; the tongue moves in response to the jaw's motion to maintain a stable constriction (Iskarous, Shadle & Proctor, 2011). Also, a cross-sectional tongue groove forms, which may help channel airflow toward the incisors, which in turn become engaged in noise generation as the jaw rises (Stone, Faber, Raphael & Shawker, 1992).
- The articulatory kinematics of /ʃ/ have not been studied in as much detail, but the difference in lingual targets may engender differences in the jaw and tongue dynamics across the two sibilants.

Peak ERB of sibilants at frication midpoint

- Due to differences in front cavity size and shape, in both English adults' productions of /s/ have higher resonances, relative to /ʃ/.
- English-acquiring children initially produce /s, ʃ/ as a single category, whose centroid value at frication midpoint is closer to adults' /s/ than /ʃ/ (Li, 2012). This same developmental pattern is exhibited in the peak ERB of the children's productions (Figure 1).

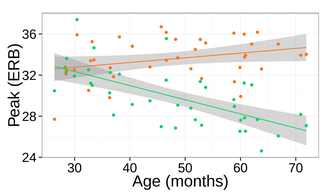


Figure 1: Development of /s/–/ʃ/ contrast in terms of midpoint peak ERB.

Previous work on the spectral kinematics of sibilants

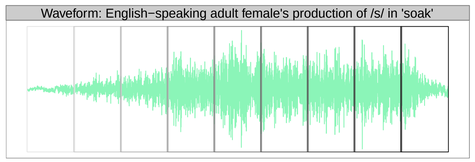
- Iskarous *et al.* (2011) approximated the centroid trajectory of adults' productions of /s/. Growth-curve models fit to these trajectories indicated that the centroid of /s/ follows a convex downward, increasing trajectory.
- Koenig, Shadle, Preston & Mooshammer (2013) measured the "development of sibilance" in adolescents' productions of /s/ by determining the change in the relative distribution of energy across low- and mid-frequency spectral bands. They found that from frication onset to frication midpoint, the energy concentration shifts from the low- to the mid-frequency range.

Method

Data collection & annotation

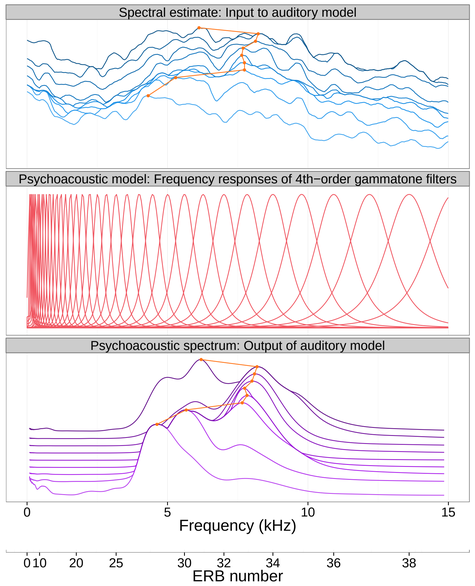
- Participants were native English-speaking adults ($N = 20$), and two- ($N = 8$), three- ($N = 14$), four- ($N = 18$), and five-year-old ($N = 19$) children who were acquiring English natively.
- Productions of /s/ or /ʃ/ in word-initial, pre-vocalic position of real English words were elicited during a audio-prompted, picture-naming task (as part of the παιδολογος project).
- Each production was transcribed, and only phonemically correct tokens were analyzed acoustically.

Computation of peak ERB trajectory



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 notched-noise 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.
- For plotting, peak ERB trajectories (orange path) were normalized by subtracting the speaker's mean peak ERB of /ʃ/. For growth curve analysis, no normalization was applied.

Growth Curve Analysis of Peak ERB Trajectories

Model specification

- `PeakERB ~ ((LinearTime + QuadraticTime) * Consonant) + ((LinearTime + QuadraticTime) | Subject:Consonant)`

Adults' productions: Interactions were significant.

- /ʃ/ increased less overall (LinearTime × Consonant: -0.8011; $se = 0.2667$; $p = 0.0027$)], and had less downward curvature (QuadraticTime × Consonant: 1.0198; $se = 0.4018$; $p = 0.0112$).

Children's productions: Interactions were almost never significant.

- LinearTime × Consonant interaction was not significant for any of the children's age groups.
- QuadraticTime × Consonant interaction across age groups:

Age	Est.	SE	*p*
2	0.2353	1.3098	0.8575
3	0.7556	0.6114	0.2165
4	0.9468	0.5674	0.095
5	1.9082	0.4566	**<1e-04**

Results: Acquisition of Peak ERB Trajectories

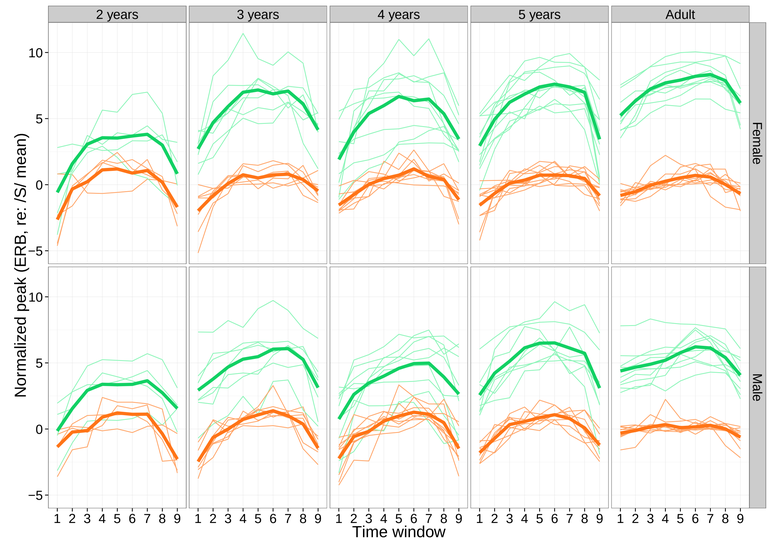


Figure 2: Individual participants' mean peak ERB trajectories for /s/ (light green) and /ʃ/ (light orange), normalized relative to their respective mean peak ERB value for /ʃ/. The mean trajectory of each AGE × GENDER demographic is overlaid as the thick, dark path.

Discussion

Adults do not produce /s/ and /ʃ/ with comparable peak ERB trajectories.

- The peak ERB trajectory of /s/ followed a convex downward, increasing trajectory, while that of /ʃ/ remained relatively flat.
- These differences in the linear and quadratic properties of their /s/ and /ʃ/ productions suggest that adults use spectral kinematic properties to differentiate these sibilants.

The ability to distinguish /s/ and /ʃ/ in terms of their spectral kinematic properties is not native in young children, but the older children tend toward an adult-like capacity to differentiate sibilants in terms of their peak ERB trajectory-curvature.

- The interaction between Consonant and either Time factor was not significant in the two-, three-, and four-year-olds.
- However, the size of the QuadraticTime × Consonant interaction increased monotonically with age, and was significant in the five-year-olds.

The ability to produce adult-like spectral kinematic patterns develops later than either the ability to produce intelligible sibilant tokens or the ability to differentiate sibilants in terms of their gross spectral features.

- All tokens analyzed were judged to be phonemically correct by a trained phonetician.
- In each age group, the main effect of Consonant was significant, suggesting that the children differentiate /s/ and /ʃ/ in terms of their mean peak ERB.
- Since the adult-like spectral kinematics develop later in childhood, they may be of benefit to studies that investigate the fine-grained, sub-phonemic properties of the /s/–/ʃ/ contrast.

Acknowledgments

- Work supported by NIDCD grant R01–02932 to Jan Edwards and by an Ohio State Graduate School Fellowship to the author.