INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

A data-driven approach for perceptually validated acoustic features for children's sibilant fricative productions

Patrick F. Reidy[1], *Mary E. Beckman*[2], *Jan Edwards*[3], *Benjamin Munson*[4]

[1]Callier Center for Communication Disorders, University of Texas at Dallas, USA
[2]Department of Linguistics, Ohio State University, USA
[3]Department of Hearing and Speech Sciences, University of Maryland, USA
[4]Department of Speech-Language-Hearing Sciences, University of Minnesota, USA

reidy@utdallas.edu, beckman.2@osu.edu, edwards@umd.edu, munso005@umn.edu

Abstract

Both perceptual and acoustic studies of children's speech independently suggest that phonological contrasts are continuously refined during acquisition. This paper considers two traditional acoustic features for the 's'-vs.-'sh' contrast (centroid and peak frequencies) and a novel feature learned from data, evaluating these features relative to perceptual ratings of children's productions.

Productions of sibilant fricatives were elicited from 16 adults and 69 preschool children. A second group of adults rated the children's productions on a visual analog scale (VAS). Each production was rated by multiple listeners; mean VAS score for each production was used as its perceptual goodness rating. For each production from the repetition task, a psychoacoustic spectrum was estimated by passing it through a filter bank that modeled the auditory periphery. From these spectra centroid and peak frequencies were computed, two traditional features for a sibilant fricative's place of articulation. A novel acoustic measure was derived by inputting the spectra to a graph-based dimensionality-reduction algorithm.

Simple regression analyses indicated that a greater amount of variance in the VAS scores was explained by the novel feature (adjusted $R^2 = 0.569$) than by either centroid (adjusted $R^2 = 0.468$) or peak frequency (adjusted $R^2 = 0.254$).
Index Terms: phonological acquisition, sibilant fricatives, Laplacian eigenmaps

1. Introduction

A large body of research suggests that phonological development is an extended and gradual process. A consistent finding in support of this characterization is that younger children's productions are both less accurate and more variable (less consistent) than older children's productions. Accuracy and variability have been measured in multiple ways. For example, accuracy traditionally is assessed using phonetic transcription by trained observers; these transcriptions can then be used to assess variability, by counting the number of different symbols that are used to transcribe multiple productions of the same target sound (e.g., [1], [2], [3]). Accuracy can also be measured in a more granular token-by-token basis by presenting multiple listeners with a word (or an excised syllable, etc.) containing a target sound and asking them to say whether the target sound was produced correctly or to rate the goodness of the production on a continuous scale (e.g., [4], [5]). For pairs of sounds in contrast, accuracy and variability can be measured using a continuous visual analog scale (VAS) between endpoints identified with the most accurate renditions of the two contrasting sounds (e.g., [6], [7]). Finally, accuracy and variability of sounds in contrast can also be assessed by measuring known acoustic features that distinguish the sounds. For example, many researchers have measured voice onset time to assess the accuracy of young children's productions of voiced versus voiceless stops (e.g., [8], [9], [10], [11], [12]), or have measured centroid frequency to assess young children's productions of sibilant fricatives (e.g., [13], [14]).

There is also some research relating these different ways of assessing accuracy and variability. For example, Munson and Urberg Carlson used the centroid frequency in English-acquiring children's productions of /s/ and /ʃ/ to compare and evaluate different methods for eliciting more granular perceptual ratings of accuracy of these sibilant fricative productions [15]. Li and colleagues used a more granular measure of accuracy (proportion of listeners who deemed the production to be a correct /s/ or a correct /ʃ/ in two different trials) to explore differences in the acoustic cues for the sibilant fricative contrast in English versus Japanese [4].

In this paper, we extend the line of research that has deployed acoustic features as an instrument for characterizing phonological development in children; however, we consider the selection of acoustic features as a process of dimensionality reduction, by which a high-dimensional representation (e.g., a spectrum) is mapped onto a low-dimensional representation (e.g., a small set of statistics computed from a spectrum). From this perspective, traditional acoustic features such as centroid frequency are determined by the researcher *a priori* with little or no consideration of the distribution or structure of the observed high-dimensional data. We propose a data-driven approach for learning acoustic features for the /s/-vs.-/ʃ/ contrast, based on a graph-theoretic algorithm (Laplacian eigenmaps; [16], [17]). We then evaluate the learned acoustic feature and two traditional features relative to perceptual ratings of preschool children's productions of /s/ and /ʃ/.

2. Method

2.1. Speech production task

The production data are part of a larger longitudinal study on relationships among speech production, speech perception, vocabulary growth, and phonological awareness. At each of three test waves, word productions were elicited in a picture-prompted word-repetition task [18]. Lists of age-appropriate target words were designed to elicit multiple tokens of /s, ʃ, t, k/ and two or three other consonants (as appropriate for the age) in word-initial position in a variety of vowel contexts. The lists for the older two test waves were also used to elicit words from

http://dx.doi.org/10.21437/Interspeech.2017-1607

adults, to evaluate pronunciation norms for the local dialect region. For the current study, word-initial voiceless sibilants /s/ and /ʃ/ were extracted from recordings of 69 of the children at the first test wave (63 female, 36 male) and 16 of the adults (10 female, 6 male). All were monolingual native speakers of American English recruited from two cities in the Northern Midwestern region of the U.S. The children ranged in age between 28 and 39 months (mean 32.9 months). The adults ranged in age between 20 and 22 years (mean 20.6 years).

The task was administered in a sound booth or a quiet room in one session (for children at any test wave) or in two sessions (for two stimulus lists for the adults). Each session was recorded using a high-quality boom microphone and stored digitally at 44.1 kHz. Each recording was annotated by a team of research assistants in two stages using customized Praat [19] scripts. At the first-stage, an annotator listened to an entire recorded session, marking off intervals corresponding to the subject's responses to the succession of stimuli and marking each interval either as an on-task production of the target word or as an off-task response (such as comment on the picture or a refusal to respond). At the second stage, an annotator listened to each on-task production for trials where the stimulus was an /s/- or /ʃ/-initial word, and identified the production of the initial consonant as a sibilant or as some other sound. If she identified the production as a sibilant, she also then marked off the boundaries for the frication noise and identified the place of articulation, choosing from five transcription categories: (1) "[s]" (i.e., a clear, unambiguous [s]), (2) "[s]:[ʃ]" (i.e., intermediate between the two sibilants but closer to [s]), (3) "[ʃ]:[s]" (intermediate but closer to [ʃ]), (4) "[ʃ]" (i.e., a clear [ʃ]), and (5) "other" (i.e., not on the [s]-to-[ʃ] continuum). The adult participants produced a total of 511 /s/ targets and 480 /ʃ/ targets as sibilant fricatives (1 item was mispronounced). The children produced a total of 896 (out of 1104) /s/ targets and 874 (out of 1104) /ʃ/ targets as sibilant fricatives that were transcribed with a place of articulation along the [s]-to[ʃ] continuum.

2.2. Perceptual rating task

The 1770 sibilant fricative productions by the children were used as the basis for stimuli in a perceptual rating task, in which adult listeners rated the /s/- or /ʃ/-likeness of each production. Productions were screened to ensure that they were free of background noise, leaving 1522 productions. From the recording of each whole-word production, the initial CV sequence was extracted, beginning 5 ms prior to the onset of sibilant frication and ending 150 ms after the onset of voicing for the vowel. (Presenting only the initial CV was intended to minimize the possibility that listeners would use lexical expectations when rating the fricative.)

Stimuli were then grouped into 4 disjoint sets of between 400 and 500 items, for experimental sessions that could be completed in less than 30 minutes, as in [15]. Seventy listeners were each administered one of the stimuli sets. All listeners were native, monolingual speakers of English between the ages of 18 and 50 years, who reported no current or previous speech, language, or hearing disorder.

On each trial, the listener saw a double-headed arrow anchored by the text "the 's' sound" at one end and "the 'sh' sound" at the other. The stimulus was played once, and the listener was asked to rate where the initial consonant fell on this visual analog scale (VAS) from an ideal /s/ to an ideal /ʃ/ by clicking at an appropriate location along the arrow. The click location in pixels was logged automatically.

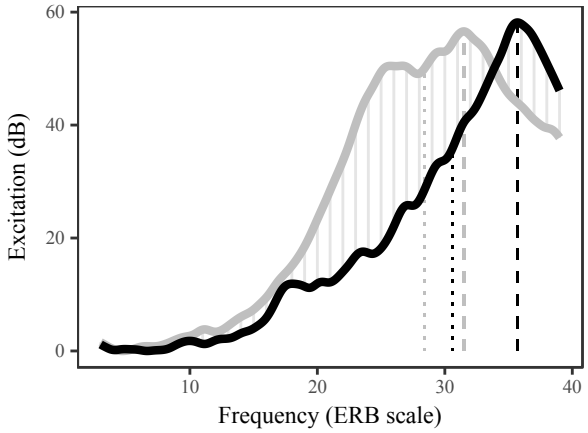

Figure 1: *Examples of psychoacoustic spectra computed from the word-initial fricatives from an adult female's productions of 'sister' (black) and 'shoes' (gray). The centroid and peak frequencies of each spectrum are indicated by dotted and dashed vertical lines, respectively. The light gray vertical segments indicate the divergence between the two spectra.*

Listeners were given no explicit instructions on what criteria they should use to judge the fricative. They were encouraged to use their 'gut instinct.' Each experiment took approximately 30 minutes to complete. A total of 37,002 ratings were elicited, across all listeners. Each stimulus item was presented to and rated by at least 10 listeners.

2.3. Acoustic analyses

Acoustic analyses were performed on all the children's productions that were used as stimuli in the VAS experiment and all the adults' productions. For each sibilant fricative production, the middle 50% of frication was extracted with a rectangular window, and then an eighth-order multitaper spectrum [20] was computed (time-bandwidth parameter $nW = 4$). A psychoacoustic spectrum was then computed by passing the multitaper spectrum through a bank of 361 gammatone filters (see also [21], [22]). The filters' center frequencies were spaced evenly between 3.0 and 39.0 along the ERB number scale, which models the auditory periphery's logarithmic frequency compression [23], [24]. The bandwidth of each filter was proportional to its center frequency, which models the auditory periphery's differential frequency resolution [25],[26]. The output of each channel in the filter bank was summed and associated to that filter's center frequency, representing the pattern of excitation induced in auditory filters (see Figure 1).

From each psychoacoustic spectrum, two traditional acoustic features for the /s/-vs./ʃ/ contrast were computed: centroid and peak frequency. Given a psychoacoustic spectrum x whose value at frequency f is denoted by x_f,

$$\text{centroid}(x) = \sum_f \frac{x_f \cdot f}{\sum_f x_f} \tag{1}$$

and

$$\text{peak}(x) = \arg\max_f \, x_f. \tag{2}$$

Both of these features index the central location of the energy distribution across the frequency scale. Several studies have reported that /s/ has greater centroid and peak frequency values

than /ʃ/ (e.g., [27]). This spectral difference is due to an articulatory difference: for /s/, place of articulation is relatively more anterior, consequently the volume of the oral cavity anterior to the constriction is relatively smaller.

A novel acoustic feature was derived by inputting the psychoacoustic spectra to a graph-based dimensionality-reduction algorithm (Laplacian eigenmaps; [16], [17]). The high-level description of this approach is that both the similarities between tokens in each talker's production space and the correspondences between different talkers' production spaces, are represented in a weighted graph, which is then projected into a low-dimensional space wherein the productions can be directly compared. More concretely, the vertices in the weighted graph represented the psychoacoustic spectra. Vertices i and j were adjacent (i.e., connected by an edge) only if (a) the corresponding excitation patterns x^i and x^j were produced by the same talker, or (b) x^i and x^j were productions of the same target word, and x^i or x^j was produced by an adult. The weight $w(i,j)$ on the edge connecting adjacent vertices i and j was

$$w(i,j) = e^{-\left(\mathcal{D}_{\mathrm{KL}}(x^i \| x^j) + \mathcal{D}_{\mathrm{KL}}(x^j \| x^i)\right)}, \tag{3}$$

where $\mathcal{D}_{\mathrm{KL}}(x^i \| x^j)$ is the Kullback-Leibler divergence from (normalized) psychoacoustic spectrum x^j to x^i

$$\mathcal{D}_{\mathrm{KL}}(x^i \| x^j) = \sum_f x_f^i \log \frac{x_f^i}{x_f^j}. \tag{4}$$

The weighted graph thus had the following properties: for a given talker, all productions were connected to each other; between any two adults, productions of the same target word were connected; between a child and an adult, productions of the same target word were connected; between any two children, no productions were connected. This particular graph structure is motivated by the facts that it represents the organization of each intra-speaker production-space and that it puts each child's production-space in correspondence with the community-norm set by the adults' production-spaces.

The graph was represented as an adjacency matrix A, such that its value on row i and column j is $A_{i,j} = w(i,j)$ (see equation (3)) if vertices i and j are adjacent in the graph, and 0 otherwise. The embedding into a low-dimensional space was found by solving the generalized eigenvalue problem

$$L\gamma = \lambda D\gamma, \tag{5}$$

where D is the diagonal matrix whose entries denote the degree of each vertex $D_{i,i} = \sum_j A_{i,j}$; and L is the Laplacian matrix $L = D - A$. A one-dimensional representation for the adults' and children's productions is given by the eigenvector γ_1 that corresponds to the least non-zero eigenvalue λ_1.

3. Results

3.1. VAS ratings

Each VAS rating in pixels was range-normalized to the interval $[0, 1]$, where 0 indicates a rating closest to the text "the 's' sound" and 1 indicates a rating closest to "the 'sh' sound." These range-normalized ratings were transformed under the empirical logit function (with adjustment 0.001) and then entered as the dependent variable in two sets of mixed-effects regression models, to determine whether ratings differed as a function of the transcription category and of target consonant, as suggested by the distribution of mean ratings shown in Figure 2. One set of models compared the ratings for stimuli that were transcribed as some kind of [s] (left half of Table 1), corresponding to the data plotted in the top half of the figure. The other set compared the ratings for stimuli that were transcribed as some kind of [ʃ] (right half of Table 1), corresponding to the data plotted in the bottom half of the figure.

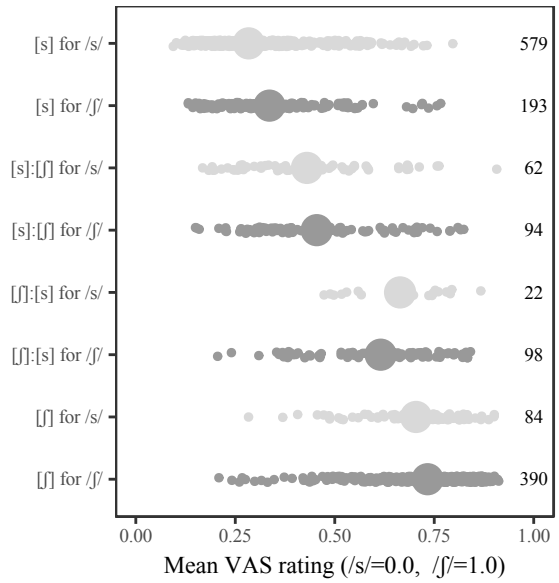

Figure 2: *Distribution of mean range-normalized VAS ratings, separated by transcription and target category. Each of the 8 large dots shows the mean value for the group of means on which it is overlaid, and each number to the right shows the number of stimuli in that group.*

Table 1: *Summary of significant fixed effects in the final model for the two sets of mixed-effects regression models.*

Transcribed as [s] or [s]:[ʃ]			**Transcribed as [ʃ] or [ʃ]:[s]**		
Coeff.	$\hat{\beta}$	t	**Coeff.**	$\hat{\beta}$	t
[s], /s/	−1.08	−14.6	[ʃ], /ʃ/	1.04	9.6
[s]:[ʃ]	0.63	17.3	[ʃ]:[s]	−0.48	−13.5
target /ʃ/	0.24	9.4	target /s/	–	–
[s]:[ʃ], /ʃ/	−0.23	−4.6	[ʃ]:[s], /s/	0.23	2.8

Adopting the step-up procedures suggested by [28], each series began with a base model that had no fixed effects and only random intercepts for talker (i.e., the subject who produced the stimulus) and for listener (the subject in the perception experiment that provided that rating). Fixed effects were then added one at a time and evaluated using the likelihood ratio test. Table 1 shows the estimated coefficients and t-values from the two final models for those effects that were shown to be significant improvements over the next simpler model.

In both series of models, there was a significant main effect of transcription category, indicating that ratings for tokens that were transcribed as "clearly [s]" (or as "clearly [ʃ]") were closer to 0 (or closer to 1) than ratings for tokens that were transcribed as intermediate (compare rows 1 and 2 of Table 1). For the first set of models (stimuli in the top half of Figure 2), there was also a significant effect of target, meaning that substitutions of (clear or intermediate) [s] for /ʃ/ were rated to be not as /s/-

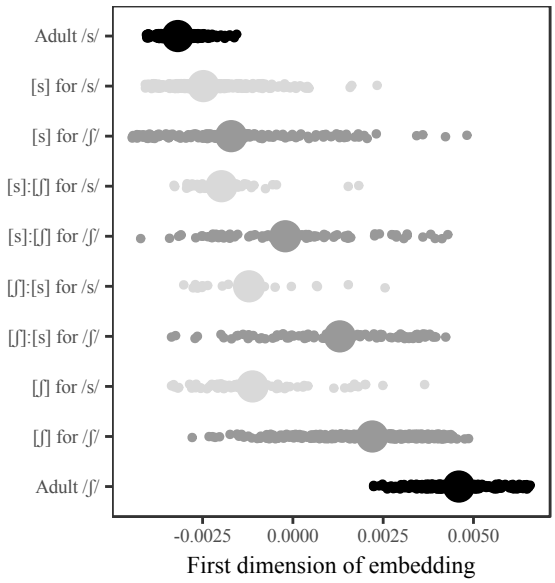

Figure 3: *The first dimension of the image of sibilant fricative productions under the Laplacian-eigenvector embedding, separated by target category for the adults (top and bottom rows) and by target and transcription category for the children (other rows). The large dots indicate the means of the subsamples.*

like as "accurate" productions of (clear or intermediate) [s] for target /s/ (see the values in the left half of row 3 of the table). Finally, for both analyses, there was a significant interaction (row 4 of the table). For the ratings of stimuli in the top half of Figure 2, this means that the effect of target on the rating was stronger for the tokens that were transcribed as clearly [s]. For the analysis of the stimuli in the bottom half of the figure, this means that ratings for stimuli that had been transcribed as more or less clearly "accurate" productions of [ʃ] for target /ʃ/ were more different from each other than the two types of transcribed substitution of [ʃ] for target /s/.

3.2. Acoustic features

Figure 3 shows the first dimension γ_1 of the Laplacian-eigenvector embedding. It is clear that γ_1 provides a feature-space within which the adults' productions of target /s/ and /ʃ/ are linearly separable independent of speaker. Conversely, the γ_1 values for the children's productions are not linearly separable by target consonant. Hence, the Laplacian-eigenvector embedding strikes a balance between separating adult community-norm productions of the target fricatives, while also preserving the greater amount of variation in the children's productions.

To evaluate how well each acoustic feature predicted the perceptual ratings, the mean proportional VAS score for each of the children's productions was logit-transformed and regressed against a single predictor: peak frequency, centroid frequency, or γ_1. Because these acoustic features had different mean values and variances, each feature was centered and scaled by its standard deviation before being entered as a predictor variable. Table 2 lists the fitted slope coefficients and measures of goodness of fit for the three models. In each model, the acoustic feature significantly predicted logit-transformed VAS score ($|t| \geq 22.76, p < 0.001$). The learned feature γ_1 better predicted the perceptual ratings, increasing adjusted R^2 by 10.1%.

Table 2: *Estimates ($\hat{\beta}$) and t-statistics for acoustic-feature coefficients; residual standard error (SE_r), adjusted R^2, and Akaike information criteria (AIC) for associated linear models.*

	$\hat{\beta}$	t	SE_r	R^2	AIC
peak	-0.555	-22.76	0.950	0.254	4163.9
centroid	-0.752	-36.57	0.802	0.468	3649.9
$\boldsymbol{\gamma_1}$	0.829	44.78	0.722	0.569	3330.0

4. Conclusions

This paper presented a method for learning acoustic features for characterizing the /s/-vs.-/ʃ/ contrast in young children. We evaluated this feature and two traditional features, in terms of how well they predicted adults' VAS ratings of the children's productions, which had not been used to supervise the mapping that yields the learned feature. Relative to centroid and peak frequency, γ_1 provided a better fit to the adults' VAS ratings.

While γ_1 outperformed the traditional features in predicting the VAS ratings, it left more than 40% of variance unexplained. The distributions for target /ʃ/ and target /s/ in Figures 2 and 3 suggest an explanation. In Figure 2, the mean VAS ratings for both targets cover the entire range from 0 to 1, with ratings for [s]-for-/ʃ/ substitutions falling in the same region as ratings for correct /s/, and with ratings for [ʃ]-for-/s/ substitutions falling in the same region as ratings for correct /ʃ/. By contrast, in Figure 3, only the values for productions of target /ʃ/ cover the entire range of γ_1 values between the adults' categories; the values for productions of target /s/ fall closer to adults' /s/ regardless of transcription category. This difference in distribution of γ_1 values for the two consonants accords with previous research. For example, the distribution of centroid values by age in a cross-language study by Li [29] suggests that young English-acquiring children's sibilant fricative productions begin as an "undifferentiated lingual gesture" [30] that is acoustically more similar to adults' /s/ than to adults' /ʃ/. The observed frequencies for the transcription categories is also in keeping with this suggested developmental path. That is, there are 287 substitutions of clear or intermediate [s] for /ʃ/ as compared to 106 substitutions of clear or intermediate [ʃ] for /s/. The distribution of mean VAS ratings across the transcription categories in Figure 2, then, might indicate adults' expectations about which sibilant will be mastered first, with greater tolerance for deviation from pronunciation norms for /ʃ/.

Because the approach described herein maps spectra to a low-dimensional representation, it is expected that the procedure would successfully characterize other sets of phonemes that differ primarily in terms of their place of articulation, which is generally reflected in spectral shape.

5. Acknowledgements

Work described in this paper was supported by NIH grant DC02932 to Edwards, Beckman, and Munson. We thank the Learning to Talk teams at UW-Madison and UMN, especially Hannele Nicholson, Bianca Schroeder, Rebecca Hatch, and Clare Kramer, for their work in recruiting and testing subjects in the production task, in annotating the recordings, and in recruiting and running subjects in the listening task.

6. References

[1] K. W. Kenney and E. M. Prather, "Articulation development in preschool children: Consistency of productions," *Journal of Speech and Hearing Research*, vol. 29, no. 1, pp. 29–36, March 1986.

[2] L. Burt, A. Holm, and B. Dodd, "Phonological awareness skills of 4-year-old British children: An assessment and developmental data," *International Journal of Language and Communication Disorders*, vol. 34, no. 3, pp. 311–335, July 1999.

[3] H.-Y. Kim and S. Ha, "Articulatory variability in 24-to 36-month-old typically developing children [in Korean]," *Communication Sciences and Disorders*, vol. 21, no. 2, pp. 333–342, June 2016. [Online]. Available: https://www.e-sciencecentral.org/articles/SC000016320

[4] F. Li, B. Munson, J. Edwards, K. Yoneyama, and K. Hall, "Language specificity in the perception of voiceless sibilant fricatives in japanese and english: Implications for cross-language differences in speech-sound development," *Journal of the Acoustical Society of America*, vol. 129, no. 2, pp. 999–1011, February 2011.

[5] T. McAllister Byun, P. Halpin, and D. Harel, "Crowdsourcing for gradient ratings of child speech: Comparing three methods of response aggregation," in *Proceedings of the 18th International Congress of Phonetic Sciences*, T. S. C. for ICPhS 2015, Ed., no. 935. Glasgow, UK: University of Glasgow, 2015. [Online]. Available: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0935.pdf

[6] B. Munson, J. Edwards, S. K. Schellinger, M. E. Beckman, and M. K. Meyer, "Deconstructing phonetic transcription: Covert contrast, perceptual bias, and an extraterrestrial view of *Vox Humana*," *Clinical Linguistics and Phonetics*, vol. 24, pp. 245–260, 2010.

[7] S. K. Schellinger, B. Munson, and J. Edwards, "Gradient perception of children's productions of /s/ and /θ/: A comparative study of rating methods," *Clinical Linguistics and Phonetics*, vol. 31, no. 1, pp. 80–103, 2017.

[8] D. Kewley-Port and M. S. Preston, "Early apical stop production: A voice onset time analysis," *Journal of Phonetics*, vol. 2, pp. 195–210, 1974.

[9] M. A. Macken and D. Barton, "The acquisition of the voicing contrast in English: a study of voice onset time in word-initial stop consonants," *Journal of Child Language*, vol. 7, pp. 41–74, 1980.

[10] J. Scobbie, F. Gibbon, W. J. Hardcastle, and P. Fletcher, "Covert contrasts as a stage in the acquisition of phonetics and phonology," in *Papers in laboratory phonology V: Language acquisition and the lexicon*, M. Broe and J. Pierrehumbert, Eds. Cambridge, U. K.: Cambridge University Press, 2000, pp. 194–207.

[11] E. J. Kong, M. E. Beckman, and J. R. Edwards, "Voice onset time is necessary but not always sufficient to describe acquisition of voiced stops: The cases of Greek and Japanese," *Journal of Phonetics*, vol. 40, no. 6, pp. 725–744, November 2012.

[12] E. R. Hitchcock and L. L. Koenig, "The effects of data reduction in determining the schedule of voicing acquisition in young children," *Journal of Speech, Language, and Hearing Research*, vol. 56, no. 2, pp. 441–457, April 2013.

[13] S. R. Baum and J. C. McNutt, "An acoustic analysis of frontal misarticulation of /s/ in children," *Journal of Phonetics*, vol. 18, no. 1, pp. 51–63, 1990.

[14] F. Li, J. Edwards, and M. E. Beckman, "Contrast and covert contrast: The phonetic development of voiceless sibilant fricatives in English and Japanese toddlers," *Journal of Phonetics*, vol. 37, pp. 111–124, 2009.

[15] B. Munson and K. Urberg Carlson, "An exploration of methods for rating children's production of sibilant fricatives," *Speech, Language, and Hearing*, vol. 19, pp. 36–45, 2016.

[16] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Advances in Neural Information Processing Systems 14 (NIPS 2001)*, T. G. Dietterich, S. Becker, and Z. Ghahramani, Eds., 2001, pp. 585–591. [Online]. Available: http://papers.nips.cc/paper/1961-laplacian-eigenmaps-and-spectral-techniques-for-embedding-and-clustering.pdf

[17] ——, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, June 2003.

[18] J. Edwards and M. E. Beckman, "Methodological questions in studying consonant acquisition," *Clinical Linguistics and Phonetics*, vol. 22, no. 12, pp. 937–956, 2008.

[19] P. Boersma, "Praat, a system for doing phonetics by computer," *Glot International*, vol. 5, no. 9/10, pp. 341–345, 2001.

[20] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proceedings of the IEEE*, vol. 70, pp. 1055–1096, 1982.

[21] A. Q. Summerfield, M. J. Nakisa, B. McCormick, S. Archbold, K. P. Gibbin, and G. M. O'Donoghue, "Use of vocalic information in the identification of /s/and /ʃ/ by children with cochlear implants," *Ear and Hearing*, vol. 23, no. 1, pp. 58–77, February 2002.

[22] P. F. Reidy, K. Kristensen, M. B. Winn, R. Y. Litovsky, and J. R. Edwards, "The acoustics of word-initial fricatives and their effect on word-level intelligibility in children with bilateral cochlear implants," *Ear and Hearing*, vol. 38, no. 1, pp. 42–56, January–February 2017.

[23] B. C. J. Moore and B. R. Glasberg, "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," *Journal of the Acoustical Society of America*, vol. 74, no. 3, pp. 750–753, September 1983.

[24] D. D. Greenwood, "A cochlear frequency-position function for several species—29 years later," *Journal of the Acoustical Society of America*, vol. 87, no. 6, pp. 2592–2605, June 1990.

[25] R. D. Patterson, "Auditory filter shapes derived with noise stimuli," *Journal of the Acoustical Society of America*, vol. 59, no. 3, pp. 640–654, March 1976.

[26] B. R. Glasberg and B. C. J. Moore, "Derivation of auditory filter shapes from notched–noise methods," *Hearing Research*, vol. 47, pp. 103–138, 1990.

[27] A. Jongman, R. Wayland, and S. Wong, "Acoustic characteristics of English fricatives," *The Journal of the Acoustical Society of America*, vol. 108, no. 3, pp. 1252–1263, 2000.

[28] H. Baayen, S. Vasishth, R. Kliegl, and D. Bates, "The cave of shadows: Addressing the human factor with generalized additive mixed models," *Journal of Memory and Language*, vol. 94, pp. 206–234, June 2017.

[29] F. Li, "Language-specific developmental differences in speech production: A cross-language acoustic study," *Child Development*, vol. 83, no. 4, pp. 1303–1315, July/August 2012.

[30] F. E. Gibbon, "Undifferentiated lingual gestures in children with articulation/phonological disorders," *Journal of Speech, Language, and Hearing Research*, vol. 42, pp. 382–397, April 1999.