Methods for analyzing time-varying spectral change in sibilant fricatives

Patrick Reidy

Waisman Center; Department of Communication Sciences & Disorders,
University of Wisconsin-Madison

This tutorial will present a general introduction to statistical methods that can be used to analyze temporal variation in the spectral properties (or, “spectral dynamics”) of turbulent sounds, such as voiceless fricatives. During the articulation of a turbulent sound, aperiodic noise sources are generated when air flowing through a narrow constriction becomes turbulent or when this turbulent airflow impinges on the incisors downstream from the constriction. While articulatory studies of turbulent sounds have reported continuous motion in the articulators responsible for generating the aperiodic noise sources (e.g., the tongue and jaw), acoustic studies of these sounds have traditionally assumed that their spectral properties are approximately stationary, representing a given spectral property of a given production with a single value, computed either from a single window of the turbulent sound, or by averaging multiple windows placed across the sound.

However, a number of recent studies have treated the spectral properties of turbulent sounds—in particular, voiceless sibilant fricatives—as quantities that vary meaningfully across the duration of the sounds. The results of these studies not only indicate that the spectral properties of turbulent sounds are dynamic, but also that differences in dynamic spectral patterns inform a host of linguistic phenomena: Iskarous, Shadle, and Proctor (2011) reported that centroid frequency varies across the duration of American English /s/. Yu (2016) reported, for Cantonese /s/, that the first four spectral moments vary across its duration and that this variation is conditioned by the following vowel context. Reidy (submitted) reported that American English /s/ and /ʃ/ contrast in terms of their peak frequency dynamics, as do Japanese /s/ and /ɕ/; furthermore, English /s/ and Japanese /s/—two cross-linguistically assimilable sounds that are produced with comparable target articulatory postures—contrast in terms of their peak frequency dynamics, but not in terms of peak frequency at fricative midpoint. Reidy (2015) reported that among children acquiring American English, 2-year-olds differentiate /s/ and /ʃ/ only in terms of static aspects of peak frequency trajectories, but 4- and 5-year-olds differentiate these fricatives in terms of both static and dynamic aspects of peak frequency. Finally, Reidy et al. (accepted) found that the intelligibility of /s/- and /ʃ/-initial words spoken by children are partly predicted by the spectral dynamics of the produced fricatives. Taken together, these studies suggest that analyzing the spectral dynamics of voiceless fricatives has the potential to improve our understanding of: anticipatory vowel-context effects, cross-linguistic differences, and the development of consonant contrasts in children.

This tutorial will present a general introduction to orthogonal polynomial growth-curve modeling—a standard method for analyzing the spectral dynamics of voiceless fricatives (see, all studies mentioned above), by which the trajectory of a spectral feature is represented as a sequence and then modeled as a linear combination of polynomial basis functions. In a such a model, it is straightforward to test whether significant differences in spectral dynamics are conditioned by linguistic factors (e.g., place-of-articulation or vowel context) by including an interaction between that factor and a basis function with a nonzero power. The core of the tutorial will introduce growth-curve models as a special case of linear mixed-effects models and will include a fully worked example demonstrating: how to fit a growth curve model; how to validate the fitted model; how to interpret the fitted model’s coefficients; and how to bootstrap prediction intervals for the fitted model. If time permits we will also discuss how to fit growth curve models in a Bayesian context or with a different class of basis functions, such as b-splines. Code examples will be given for the programming language R.

REFERENCES

Iskarous, K. Shadle, C.H. and Proctor, M.I. (2011). Articulatory-acoustic kinematics: The production of American English /s/. *J. Acoust. Soc. Am.*, 129(2):944–954.

Reidy, P.F. (2015). *The spectral dynamics of voiceless sibilant fricatives in English and Japanese*. Ph.D. dissertation, The Ohio State University, Columbus, OH. pp. 63–88.

Reidy, P.F., Kristensen, K., Winn, M.B., Litovsky, R.Y., and Edwards, J.R. (accepted). The acoustics of word-initial fricatives and their effect on word-level intelligibility in children with bilateral cochlear implants. *Ear Hearing*.

Reidy, P.F. (submitted). Spectral dynamics of sibilant fricatives are contrastive and language specific. *J. Acoust. Soc. Am.*

Yu, A.C.L. (2016). Vowel-dependent variation in Cantonese /s/ from an individual-difference perspective. *J. Acoust. Soc. Am.*, 139(4):1672–1690.