

AURORA model of formant-to-tongue inversion for didactic and clinical applications

Sam Kirkham¹, Patrycja Strycharczuk²

¹Lancaster University & ²University of Manchester

Introduction AURORA (Acoustic Understanding and Real-time Observation of Resonant Articulations) is an articulatory inversion model, which predicts two-dimensional information about tongue displacement/shape in vowels, using formant values as the input. In this paper, we outline the conceptual and computational foundations of AURORA, and illustrate potential applications for visual biofeedback in phonetics teaching and speech therapy [2].

Motivation Vowel formants are a fundamental notion in speech science. This notion sometimes needs to be explained in an accessible way to wider audiences, such as academic audiences with a linguistics background, but not specialised phonetic knowledge, or selected speech therapy clients [1, 2]. One potential way of delivering an accessible explanation involves linking the formants to the underlying lingual articulation. Our tool facilitates this explanation by providing a visual display of tongue movement in response to manipulation of F_1 and F_2 . It achieves a degree of articulatory accuracy thanks to modelling the interdependence of multiple articulatory dimensions and the interdependence of F_1 and F_2 , both of which are informed by empirical data. It achieves simplicity by providing readily interpretable and low-dimensional visual output.

Model architecture AURORA is informed by ultrasound tongue imaging and acoustic data from 40 L1 speakers of English [3]. The model was trained using a highly-interpretable and data-efficient method, with a multivariate linear model mapping the first two formants onto six articulatory parameters that model tongue shape.

Evaluation and extensions We present a qualitative evaluation of the model, focusing on selected tongue features, which shows good prediction accuracy of tongue shape. We then present two tools developed to make the model more accessible to a wider audience, a Shiny app, and software for real-time tongue biofeedback from acoustics (Figure 1). The Shiny app is intended primarily as a didactic aid that can be used in phonetics teaching, as well as in any context that requires explaining formants. The biofeedback tool has potential for use in some forms of speech therapy, e.g. Gender Affirming Voice Training [2]. Potential users include students of phonetics, linguists in fields adjacent to phonetics, as well as speech therapy practitioners and clients.

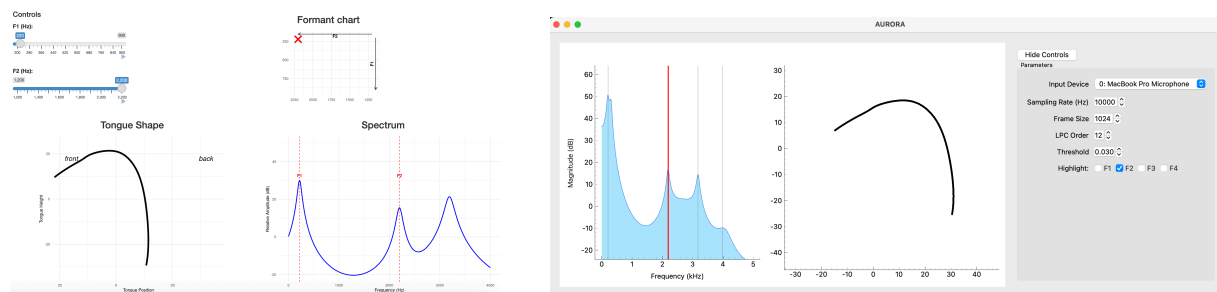


Figure 1: Left: Shiny web app, showing the controls and the predicted tongue shape output. Right: The real-time biofeedback app, showing spectral and tongue shape feedback.

References [1] D. Kawitzky and T. McAllister. “The effect of formant biofeedback on the feminization of voice in transgender women”. In: *Journal of Voice* 34 (2020). [2] T. McAllister et al. “Real-Time Resonance Biofeedback for Gender-Affirming Voice Training: Usability Testing of the TruVox Web-Based Application”. In: *Journal of Voice* (2025). [3] P. Strycharczuk and S. Kirkham. “Articulatory strategies in male and female vowel production”. In: *Journal of Speech, Language, and Hearing Research* (2025).