

DYNAMICS OF VOCAL TRACT SHAPING

TEAM: BYRD, LEAHY, LEE, NAYAK, NARAYANAN (PI)

AGENCY: NIH

DATES: 2005-2009

A. Specific Aims

The vocal tract is the universal human instrument, played with great dexterity and skill in the production of spoken language. In order to produce the elegant acoustic structure of speech, the ongoing shaping of the vocal tract must be choreographed with moment-to-moment precision, as it is this vocal tract shaping, in coordination with respiratory and laryngeal behavior, that produces the linguistically significant resonant structure of vowels and sonorant consonants and the complex aerodynamic qualities of obstruent consonants.

The **long-term goal** of this project is to wed state-of-the-art technology for imaging the vocal tract with a linguistically informed analysis of the speech tasks or goals requisite in the production of spoken language. Importantly, this includes furthering our understanding of the relations between vocal tract shaping and speech acoustics. In the past, magnetic resonance imaging (MRI) has served as a valuable tool for studying static vocal tract postures. Now, our team has developed an acquisition strategy, described in detail below, that allows for image reconstruction rates of 24 images per second, making veridical real-time movies of speech production possible for the first time without X-rays (Narayanan, Nayak, Lee, Sethy, & Byrd, in press). Data show clear real-time movements of the lips, tongue, and velum, providing exquisite information about the spatiotemporal properties of speech gestures in both the oral and pharyngeal portions of the vocal tract. *Our long-term goal is to understand which aspects of vocal tract shaping are critically controlled during speech*, both for sounds known to be complex in geometry (e.g., /r/ & sibilant fricatives) and for sounds known to be complex in their temporal structuring (e.g., /l/ & diphthongs). An understanding of vocal tract shape as a fundamentally dynamic aspect of linguistic organization will do much to add to the field's current—basically *static* (i.e. postural & fixed time-point)—approach to describing the production of speech.

The **specific aims** of this proposal are to further develop the technology and analysis platform of real-time MRI, which provides the scaffolding for the project, while pursuing speech production studies in three areas related to phonological representation: sounds thought to have sequential vocal tract constriction goals, sounds with prosodically-sensitive coordination of vocal tract constriction goals, and sounds with geometrically complex vocal tract shaping goals. Specifically, we propose to use the real-time MRI approach we've developed to understand the internal linguistic structuring of:

- *Units with Sequential Goals—Diphthongs*
Competing hypotheses of two sequential production targets versus one elaborated production target in diphthongs—e.g. [ai] (as in “I”) and [au] (as in “how”)—will be investigated by examining variability at diphthong endpoint in articulator position, constriction degree, and vocal tract shape.
- *Units with Prosodically Sensitive Goals—Multi-Gesture Segments*
Syllabic-structure and stress will be used as a probe to examine coordination within the multi-gesture segments [r], [l], and [n] to determine how the coordination of the multiple and highly overlapped constrictions is realized in space and time.
- *Units with Geometrically Complex Goals—Fricatives and Liquids*
The hypothesis, based on static vocal tract posture data for the English grooved fricative [s] and the Tamil “cupped” retroflex liquid [ʁ], that overall vocal tract shaping (as opposed to solely a point

constriction) is requisite in the production these sounds will be evaluated by examining the realization of canonical tongue shaping under conflicting coarticulatory forces.

An appropriate understanding of how these sounds are produced in space and time is fundamentally a phonological question in that it bears directly on the phonological representation of segmental units, a representation that we take to be intrinsically articulatory and dynamic, as proposed by Browman & Goldstein's framework of Articulatory Phonology (1992). To date, the specification of the phonological properties of a segment within Articulatory Phonology have included one or more constriction gestures, each with a single mid-sagittal constriction location and degree specified. Our work will augment how phonological representation in articulatory terms is understood by considering situations in which dynamic movement, coordination of multiple constrictions, or critical cross-sectional vocal tract shaping are likely to be a essential elements of phonological specification.

The project's final specific aim concerns advances in our technical approach to investigating the physical realization of phonological structure:

- *Real-Time Technology & Data Analysis*

We plan the development of more sophisticated pulse-sequence, head-neck coil, and acoustic recording approaches for real-time MRI, in tandem with data-driven analysis approaches suitable for distilling the high-dimensional information provided by real-time MRI. Ultimately, through integrated analysis and modeling of dynamic image data and speech acoustic data, we aim to gain an improved understanding of the fine details of dynamic articulatory-acoustic relations.

We feel that it is no exaggeration to say that the advent of real-time MRI for speech has the potential to literally change the way speech production research is conducted in the field.