Does Articulatory Setting Provide Some Mechanical Advantage For Speech Motor Action?

Vikram Ramanarayanan, Adam Lammert, Louis Goldstein, and Shrikanth Narayanan

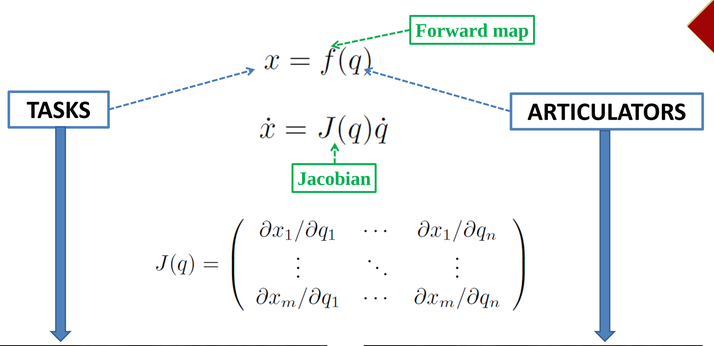
University of Southern California, Los Angeles, CA

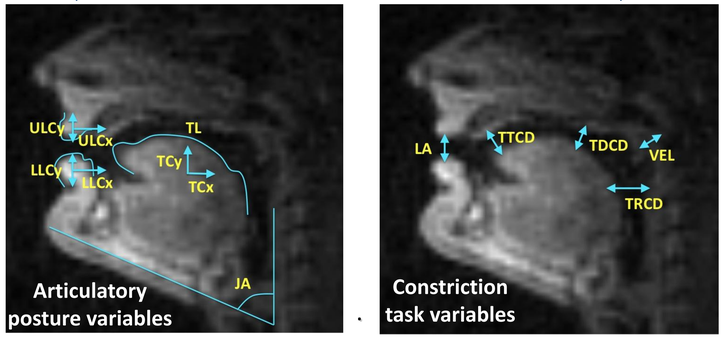
[Supported by NIH]

http://sail.usc.edu/span

DIFFERENTIAL KINEMATICS & MECHANICAL ADVANTAGE

The relationship between **low-level articulators and high-level tasks/goals** is commonly expressed by the direct kinematics equation, of the form:





To combine these individual speed ratios into an overall measure of mechanical advantage, the sum of squares of all Jacobian values was used:

$$MA = \sum_{i=1}^{n}\sum_{j=1}^{m} J_{i,j}^2$$

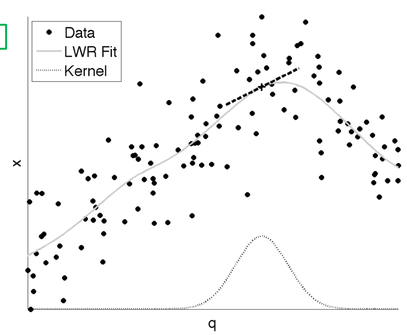
Jacobian can be estimated to a high degree of accuracy in data-driven fashion using Locally-Weighted Linear Regression (LWR) [3]

$$x_i = \beta_i^T q_i$$
$$\beta_i = (Q^T W_i Q + \gamma I)^{-1} Q^T W_i X$$

Weighting kernel

Uses locally-defined, lower-order polynomials to approximate a global function

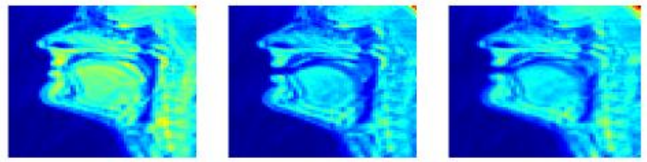
Closed-form solution with minimal parameter tuning



WHAT IS ARTICULATORY SETTING?

The set of postural configurations (which can be language-specific and/or speaker-specific) that the vocal tract articulators tend to be **deployed from** and **return to** in the process of producing fluent and natural speech

We found earlier [1] that vocal tract postures at absolute rest are *significantly different* and *more variable* as compared to those during inter-speech pauses and speech-ready positions.



DATA COLLECTION AND PROCESSING

1 Real-time MRI provides a mid-sagittal view of all supraglottal vocal tract articulators [1,2].

5 speakers of American English – read speech data.

Simultaneous MRI and noise-cancelled audio; phonetically-aligned.

2 Use phonetic alignment to divide each speaker's data into 11 categories

Vowels
- High front
- High back
- Low front
- Low Back

Consonants
- Labial
- Coronal
- Dorsal
- Approximants

Rest
Absolute rest; posture before scanner is on

Inter-speech Pause
Grammatical pauses during speech

Ready
100-200ms before subject starts speaking

3 For all extracted frames for a given speaker, task & articulatory variables computed and normalized by its range

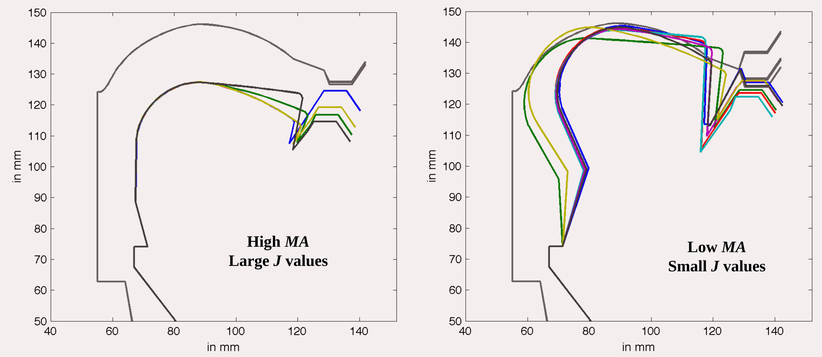
4 **Jacobians estimated using bootstrapping (N = 100 samples)**

In each bootstrap iteration, a "test" posture was randomly sampled from all the postures in that category.

The LWR model was then fit to the rest of the data (training data), which was then used to estimate a Jacobian matrix for the test posture.

OBSERVATIONS AND RESULTS

Simulation experiments using the Configurable Articulatory Synthesizer (CASY)



EXPERIMENTS: rt-MRI data

A non-parametric 2-way Friedman's test was performed:
- fixed factor = 11 linguistic categories
- random factor = 5 speakers
- N = 100 replicates – bootstrapping procedure
- Non-parametric Mann-Whitney U post-hoc tests.

Category		Eng1	Eng2	Eng3	Eng4	Eng5	Rest	Ready	Vowels HF	Vowels HB	Vowels LF	Vowels LB	Consonants Lab	Consonants Cor	Consonants Dor	Consonants App.
		Medians of SS Jacobian					Number of speakers with significant pairwise differences in median									
ISP		11.28	13.16	6.42	24.43	20.73	3	3	4	4	4	5	5	4	5	4
Rest		1.50	10.49	4.74	17.69	22.28		4	3	5	5	4	5	4	4	4
Ready		11.98	10.60	7.68	19.45	20.85			4	5	5	4	5	5	5	5
Vowel	HF	9.94	7.83	6.41	17.56	20.44				4	4	3	5	4	3	4
	HB	11.23	5.56	9.68	18.22	19.97					4	4	2	4	4	4
	LF	8.95	5.07	6.74	18.14	17.50						3	4	4	4	5
	LB	9.60	6.72	7.71	18.86	19.35							3	3	0	2
Cons.	Lab	10.97	5.39	9.87	18.89	18.68								2	3	2
	Cor	10.93	8.21	7.17	18.89	19.28									3	2
	Dor	10.00	6.62	7.26	19.33	19.53										2
	App.	10.69	7.04	8.72	19.26	19.40										

Speech-ready postures, ISPs are generally more mechanically advantageous than postures during absolute rest.(caveat: Eng5)

Case of vowel and consonant postures is not as clear

Open questions:

(1) Which task/articulatory variables?

(2) How do you best define mechanical advantage?

(3) Can we use kinematics ideas to inform phonological theories?

REFERENCES

1. Ramanarayanan V., Goldstein L., Byrd D., and Narayanan S. (2013). " An investigation of articulatory setting using real-time magnetic resonance imaging", *J. Acoust. Soc. Am.* 134 (1).
2. Narayanan S., Nayak K., Lee S., Sethy A., and Byrd D. (2004). "An approach to real-time magnetic resonance imaging for speech production," *J. Acoust. Soc. Am.* 115, 1771–1776.
3. Lammert A., Goldstein L., Narayanan S., Iskarous K. (2013). "Statistical methods for estimation of direct and differential kinematics of the vocal tract. *Speech Communication* 55(1): 147-161.