Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Speech Communication
Session 5aSCb: Production and Perception II: The Speech Segment (Poster Session)

5aSCb28. Does articulatory setting provide some mechanical advantage for speech motor action?

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

***Corresponding author's address: University of Southern California, Los Angeles, CA 90007, vramanar@usc.edu**

Articulatory setting postures adopted during speech production are examined with the goal of determining whether setting postures are more mechanically advantageous than rest positions in facilitating motion of vocal tract articulators toward task goals. Articulatory simulations using the Task Dynamics Application (TADA) suggest that setting postures afford large changes with respect to speech tasks for relatively small changes in low-level speech articulators, thus affording greater mechanical advantage as compared to absolute rest postures. This study investigates this hypothesis using real-time Magnetic Resonance Imaging (rtMRI) data of read and spontaneous speech elicited from 5 healthy speakers of American English. Frames corresponding to inter-speech pauses, speech-ready intervals and absolute rest intervals were identified and image features were automatically extracted to quantify the vocal tract postures in terms of both task-level constriction variables and articulatory variables. Locally Weighted Regression is then used to estimate the ratio of task velocities to articulator velocities (i.e., the lever or speed ratio) at postures corresponding to the different intervals of interest. Results show substantially higher speed ratios at inter-speech and ready postures as compared to absolute rest postures.

Published by the Acoustical Society of America through the American Institute of Physics

 [DOI: 10.1121/1.4800246]
Received 21 Jan 2013; published 2 Jun 2013

ARTICULATORY SETTING AND MECHANICAL ADVANTAGE

Articulatory setting (also called phonetic setting or organic basis of articulation or voice quality setting; henceforth referred to as AS) may be defined as 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 (Sweet, 1890; Honikman, 1964; Laver, 1978; Esling and Wong, 1983). AS differs during rest positions, ready positions and read inter-speech pauses and, in that order, exhibit a trend for *decreasing* variability and, thus, a possible *increasing* degree of active control by the cognitive speech planning mechanism (Ramanarayanan *et al.*, 2010, 2011). In this work, AS postures adopted during speech production are examined with the goal of determining whether setting postures can be meaningfully defined in terms of the mechanical characteristics of the speech production apparatus. Specifically, this study tests the hypothesis that AS more mechanically advantageous than absolute rest positions, facilitating rapid motion of vocal tract articulators toward task goals.

The central hypothesis of this study is related to the relationship between different descriptions of the speech production apparatus, representing various levels of the system, from muscle activations to constriction degrees/locations, and up to formant frequencies. Low-level descriptions are assumed to be closer to the articulatory substrate (i.e., the actively controlled components) while high-level variables have the potential to be task variables, used to define the desired actions of the system. An important aspect of the kinematic relationships between these variables is the *speed ratio*, defined as the ratio of task variable velocities to articulator velocities (Antonsson and Cagan, 2005; Siciliano and Khatib, 2008). Speed ratios with large numerical values are mechanically advantageous because relatively small changes in articulators can result in large changes in task space. Mechanical advantage (MA) is classically defined in terms of force amplification, rather than speed amplification. However, these definitions are the same if one assumes preservation of power from articulators to tasks (i.e., the system is ideal). Perhaps the simplest example MA is provided by a basic (i.e., Class 2) lever, which amplifies force and speed on different sides of the fulcrum according to the ratio of lengths of those sides.

With this in mind, the specific hypothesis of this study can be stated more precisely. The expectation is that postures with the highest mechanical advantage of this kind bear resemblance to articulatory setting, and postures with the lowest mechanical advantage are more extremal vocal tract postures (for example, those observed during certain vocal tract constriction actions). This study is aimed at quantitatively testing this hypothesis, using both articulatory vocal tract simulations and real speech data.

CALCULATING MECHANICAL ADVANTAGE

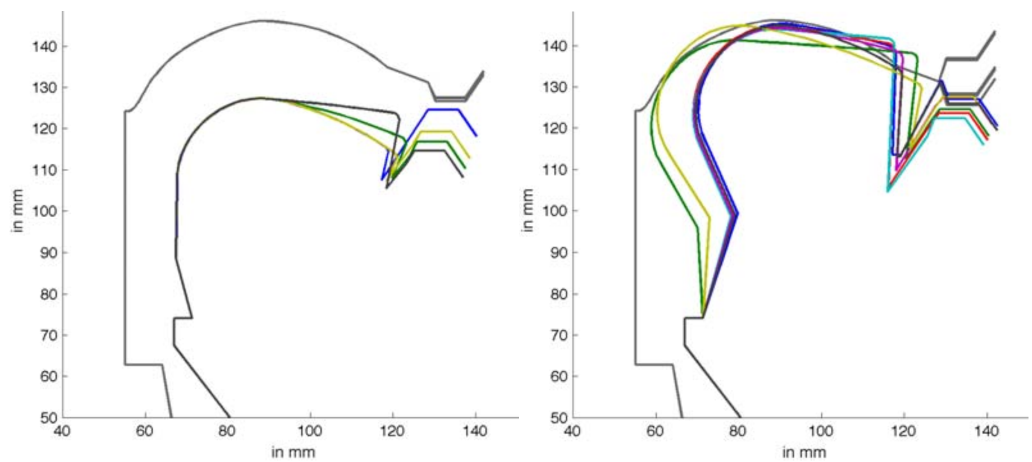
It is possible to accurately estimate kinematics of the vocal tract in data-driven fashion. It was recently shown that Locally-Weighted Linear Regression (LWR) is useful for this purpose, producing accurate estimation and having practical advantages (Lammert *et al.*, in press). LWR is a method that uses locally-defined, low-order polynomials to approximate globally nonlinear functional relationships. Training the model has a closed-form solution via the generalized least squares solution. The method also has few free parameters, making accurate training even more feasible.

The result is a Jacobian matrix, relating the velocities of each task to each articulator. Each value in the Jacobian represent a speed ratio that could be used to characterize MA in the system. To have a single measure of MA for each posture, an overall measure of

MA was simply defined as the mean of all Jacobian values.

Simulation Experiments

Confirmation of this reasoning was initially sought in kinematic models of two motor apparatuses: a multi-link planar robot arm with revolute joints and the Task Dynamics Application (TADA) (Nam *et al.*, 2004, 2006). The kinematic model at the core of TADA is the Configurable Articulatory Synthesizer (CASY) (Rubin *et al.*, 1996; Iskarous *et al.*, 2003), based on the geometrical equations of Mermelstein (1973) for describing vocal tract configurations and deriving vocal tract outlines from articulator positions. This model was recently used to compile the TADA-TIMIT database, containing complete articulatory/kinematic data for 460 sentences of English.



(A) Most Advantageous **(B)** Least Advantageous

Figure 1: Articulatory configurations corresponding to the top eight most and least mechanically advantageous postures observed in the TADA-TIMIT data set.

Experiments using Real-Time Magnetic Resonance Imaging Data

Real-time Magnetic Resonance Imaging (rtMRI) data of read and spontaneous speech was elicited from 5 healthy speakers of American English. Frames corresponding to inter-speech pauses, speech-ready intervals and absolute rest intervals were identified and image features were automatically extracted to quantify the vocal tract postures in terms of both task-level constriction variables and articulatory variables. Locally Weighted Regression was then used to estimate the ratio of task velocities to articulator velocities (i.e., the lever or speed ratio) at postures corresponding to the different intervals of interest.

Conclusions

Experiments presented here, both in simulation and with real speech data support the central hypothesis. Articulatory simulations suggest that those postures which afford the largest overall task-level changes for relatively small changes in low-level articulators are precisely those postures which bear close resemblance to setting postures. Thus, setting postures afford greater mechanical advantage as compared to absolute rest postures. In conformity with simulation experiments, results from real speech data show substantially higher speed ratios at inter-speech and ready postures as compared to absolute rest postures.

ACKNOWLEDGMENTS

REFERENCES

Antonsson, E. and Cagan, J. (**2005**). *Formal Engineering Design Synthesis* (Cambridge University Press).

Esling, J. and Wong, R. (**1983**). "Voice quality settings and the teaching of pronunciation", TESOL Quarterly **17**, 89–95.

Honikman, B. (**1964**). "Articulatory settings", In D. Abercrombie, D.B. Fry, P.A.D. MacCarthy, N.C. Scott and J.L.M. Trim (eds.), In Honour of Daniel Jones 73–84.

Iskarous, K., Goldstein, L., Whalen, D., Tiede, M., and Rubin, P. (**2003**). "Casy: The haskins configurable articulatory synthesizer", in *Proceedings of the International Congress of Phonetic Sciences*.

Lammert, A., Goldstein, L., Narayanan, S., and Iskarous, K. (**in press**). "Statistical methods for estimation of direct and differential kinematics of the vocal tract", Journal of Speech Communication .

Laver, J. (**1978**). "The concept of articulatory settings: an historical survey", Historiographia Linguistica, 5 **1**, 1–14.

Mermelstein, P. (**1973**). "Articulatory model for the study of speech production", Journal of the Acoustical Society of America **53**, 1070–1082.

Nam, H., Goldstein, L., Browman, C., Rubin, P., Proctor, M., and Saltzman, E. (**2006**). *TADA (TAsk Dynamics Application) manual*.

Nam, H., Goldstein, L., Saltzman, E., and Byrd, D. (**2004**). "Tada: An enhanced, portable task dynamics model in matlab", Journal of the Acoustical Society of America **115**, 2430–2430.

Ramanarayanan, V., Byrd, D., Goldstein, L., and Narayanan, S. (**2010**). "Investigating articulatory setting–pauses, ready position, and rest–using real-time MRI", Interspeech, Makuhari, Japan .

Ramanarayanan, V., Goldstein, L., Byrd, D., and Narayanan, S. (**2011**). "An MRI study of articulatory settings of L1 and L2 speakers of American English", International Seminar on Speech Production, Montreal, Canada .

Rubin, P., Saltzman, E., Goldstein, L., McGowan, R., Tiede, M., and Browman, C. (**1996**). "Casy and extensions to the task-dynamic model", in *Proceedings of the 1st ETRW on Speech Production Modeling, Autrans, France*.

Siciliano, B. and Khatib, O., eds. (**2008**). *Springer Handbook of Robotics* (Springer).

Sweet, H. (**1890**). *A primer of phonetics* (Clarendon Press).