Characterizing vocal tract dynamics with real-time MRI

Real-time magnetic resonance imaging (rtMRI) provides information about the dynamic shaping of the vocal tract during speech production [1]. We develop a dynamical system in the framework of Task Dynamics [2] which controls vocal tract constrictions and induces deformation of the air-tissue boundary. The system parameterizes the air-tissue boundary through factor analysis of rtMRI videos and characterizes the relation between vocal tract shape and vocal tract constrictions.

We used midsagittal rtMRI videos of the first 250 utterances of speaker F1 from the USC-TIMIT database [3]. The algorithm of [4] was used to segment the air-tissue boundary of each rtMRI video frame. A factor analysis parameterized the air-tissue boundary as a linear combination of the components of Figure 1 [5].

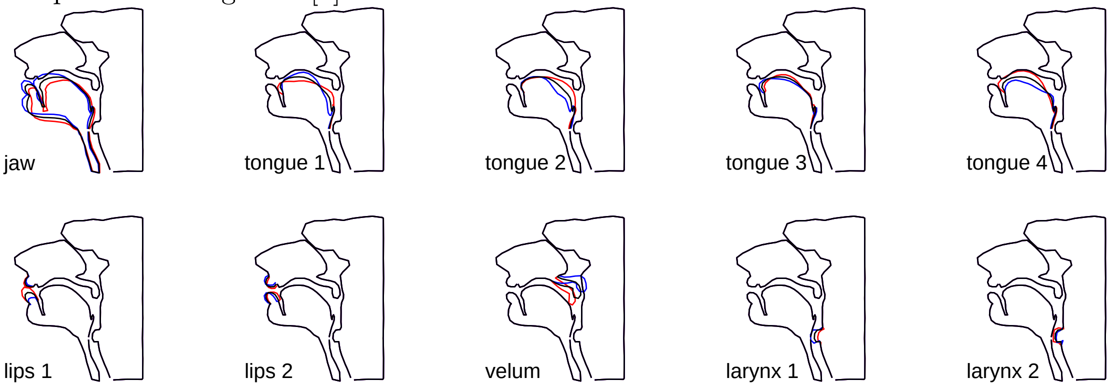

Figure 1: Components of the factor analysis

We reconstructed the air-tissue boundary in each frame as a weighted linear combination of the components (RMS error is 1.7 mm). From the reconstructed air-tissue boundary we used an algorithm adapted from [6] to measure the degree of constriction in millimeters at the bilabial, alveolar, palatal, velar, pharyngeal, and velopharyngeal port regions (Figure 2).

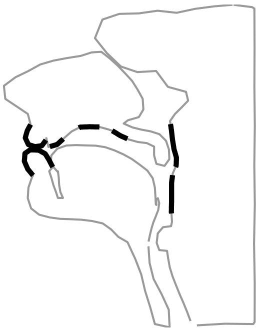

Figure 2: Places of constriction

Thus, each rtMRI video has both a path $\mathbf{w}(t) \in \mathbf{R}^{10}$ of factor weight vectors and a path $\mathbf{z}(t) \in \mathbf{R}^{6}$ of constriction degree vectors sampled at each frame. We developed an algorithm to estimate the forward map [7, 8] $\mathbf{G} : \mathbf{R}^{10} \to \mathbf{R}^{6}$ which maps weight vectors to constriction degree vectors. The algorithm computes a tree whose root node is the set of all 18130 observed weight vectors in $\mathbf{R}^{10}$. A k-means subroutine starts at the root and iteratively breaks nodes in two (i.e., $k = 2$). Children in this tree are disjoint subsets of the parent and the union of siblings is the parent. Nodes stop breaking either when a child would contain fewer than seven weight vectors (to prevent rank-deficiency in least squares estimation of $\mathbf{G}$) or when $\mathbf{G}$ maps the weight vectors of that node to constriction degree vectors in $\mathbf{R}^{6}$ approximately linearly (i.e., when $\mathbf{G}(\mathbf{w})$ estimates $\mathbf{z}$ with a mean absolute error of less than 0.24 mm). The tree has 198 terminal nodes, 87% of which are volumes of $\mathbf{R}^{10}$ in which $\mathbf{G}$ is approximately linear, and 13% of which were too small to continue breaking. Within each terminal node, the algorithm uses least squares to estimate $\mathbf{G}$, the jacobian J of $\mathbf{G}$, and the time derivative $\dot{J}$ of the jacobian.

We describe change in the vector $\mathbf{z}$ of constriction degrees over time with the differential equation $\ddot{\mathbf{z}} = -K(\mathbf{z} - \mathbf{z}_0) - B\dot{\mathbf{z}}$, where $\mathbf{z}_0$ is a vector of six constriction degree targets and K and B are 6×6 diagonal matrices of stiffness and damping coefficients, respectively [2]. For every path $\mathbf{z}(t)$ of constriction degrees which solves the differential equation, we have infinitely many paths of weight vectors which describe the deforming air-tissue boundary [9]. We find one such weight path $\mathbf{w}(t)$ by solving the differential equation $\ddot{\mathbf{w}} = J^*(-BJ\dot{\mathbf{w}} - K(\mathbf{G}(\mathbf{w}) - \mathbf{z}_0)) - J^*\dot{J}\dot{\mathbf{w}}$. This follows from the change of variables $\mathbf{z} = \mathbf{G}(\mathbf{w})$, $\dot{\mathbf{z}} = J\mathbf{w}$, $\ddot{\mathbf{z}} = J\dot{\mathbf{w}} + \dot{J}\mathbf{w}$ and from the pseudoinverse J^* of J from [2]. Figure 3 shows model predictions for how the air-tissue boundary deforms for bilabial, alveolar, palatal, velar, pharyngeal, and velopharyngeal port constrictions starting from an articulatory setting position where weights are set to their means.

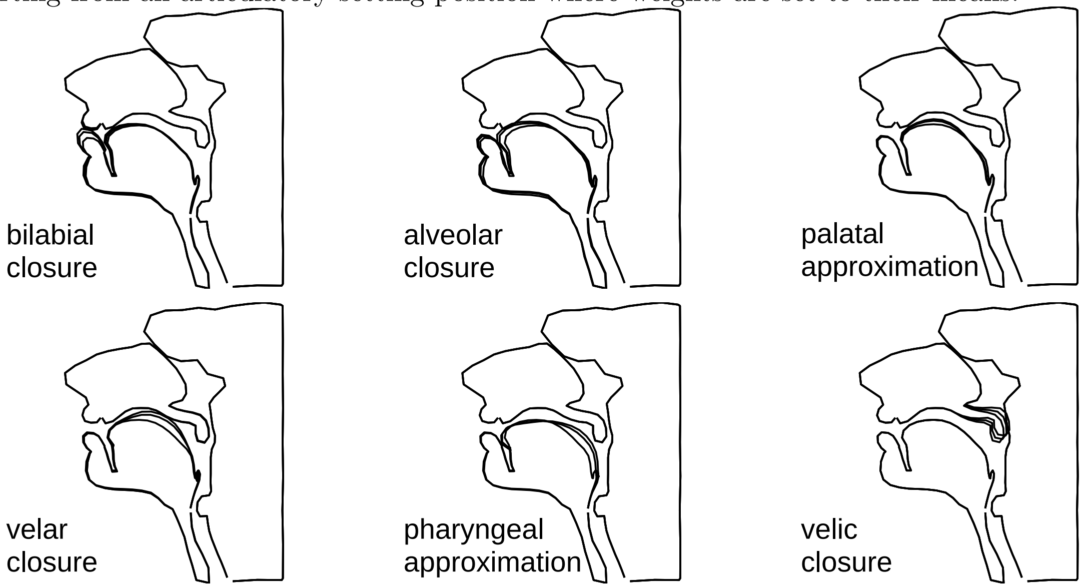

Figure 3: Air-tissue boundary deformation in response to constrictions

For each observed constriction degree vector $\mathbf{z}_i$ we approximated the corresponding weight vector $\mathbf{w}_i$ as $\hat{\mathbf{w}}_i = \mathbf{G}^*\mathbf{z}_i$ using the pseudoinverse $\mathbf{G}^*$ of $\mathbf{G}$ and reconstructed the air-tissue boundary using $\hat{\mathbf{w}}_i$. RMS error is 2.6 mm when compared against the air-tissue boundary segmented from the rtMRI video [4]. Thus, reconstruction of the air-tissue boundary from constriction degree vectors shows agreement with the observed air-tissue boundaries.

In sum, we have used rtMRI to develop a task-dynamical system which characterizes the relation between factors of vocal tract shape and vocal tract constrictions. This is the first task-dynamical model explicitly derived from data which can then be used to quantitatively assess the model. Possible applications include speaker-specific articulatory speech synthesis and characterizing inter- and intra-speaker variability. We plan to apply and test these modeling methodologies on more speakers in the USC-TIMIT database.

[1] S. Narayanan et al. “An approach to real-time magnetic resonance imaging for speech production”. In: *Journal of the acoustical society of America* 115.4 (2004). [2] E. L. Saltzman and K. G. Munhall. “A dynamical approach to gestural patterning in speech production”. In: *Ecological psychology* 1.4 (1989). [3] S. Narayanan et al. “Real-time magnetic resonance imaging and electromagnetic articulography database for speech production research (TC)”. In: *The Journal of the Acoustical Society of America* 136.3 (2014). [4] E. Bresch and S. Narayanan. “Region segmentation in the frequency domain applied to upper airway real-time magnetic resonance images”. In: *Medical Imaging, IEEE Transactions on* 28.3 (2009). [5] A. Toutios and S. S. Narayanan. “Factor Analysis of Vocal-Tract Outlines derived from Real-Time Magnetic Resonance Imaging Data”. In: *18th International Congress of Phonetic Sciences (ICPhS).* Glasgow, Scotland, UK, 2015. [6] V. Ramanarayanan et al. “An investigation of articulatory setting using real-time magnetic resonance imaging”. In: *The Journal of the Acoustical Society of America* 134.1 (2013). [7] A. Lammert et al. “Statistical methods for estimation of direct and differential kinematics of the vocal tract”. In: *Speech communication* 55.1 (2013). [8] S. Ouni and Y. Laprie. “Modeling the articulatory space using a hypercube codebook for acoustic-to-articulatory inversion”. In: *The Journal of the Acoustical Society of America* 118.1 (2005). [9] N. A. Bernstein. *The co-ordination and regulation of movements.* Pergamon Press Ltd., 1967.