USC University of Southern California SAIL

*Supported by NIH grant R01DC007124 and NSF grants 1514544, 1908865*

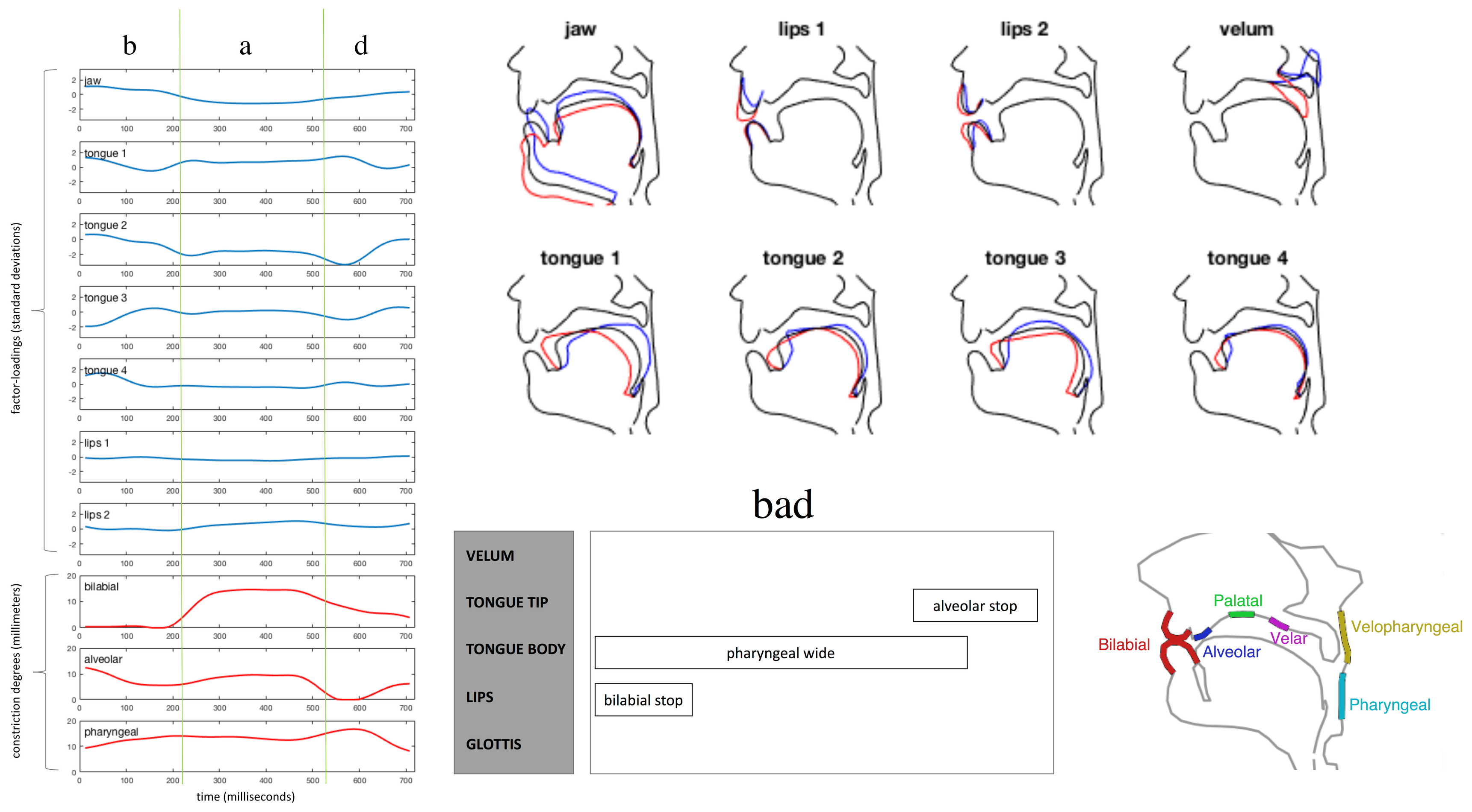
# Simulating anticipatory coarticulation in VCV utterances with a gestural articulatory synthesizer

*Asterios Toutios and Shrikanth Narayanan*

## Introduction

- **Context:** Development of a framework for articulatory speech synthesis from gestural score specifications directly informed by real-time MRI data (Alexander et al., 2019)
- Proposing a modified/enhanced gestural score
- Describing the mechanism by which such gestural score animates (a midsagittal slice of) the vocal tract
- Addressing gestural overlap
- Test case: coarticulation in /adu/, as observed by Öhman (1966)

## Articulatory Model and Forward Map




- Articulatory model developed by guided PCA approximates midsagittal vocal tract shaping by 8 control parameters
- Constriction degrees measured at 6 places of articulation
- Joint space is modeled as a union of clusters where the *forward map* from parameters **w** to constrictions **z** is approximately linear, i.e. $\mathbf{z} = \mathbf{G}(\mathbf{w}) = F * \mathbf{w} + \mathbf{z_c}$

## Discretization of Dynamical Systems

- From Saltzman and Munhall (1998):

$$\ddot{\mathbf{w}} = J^*(-BJ\dot{\mathbf{w}} - K(\mathbf{G}(\mathbf{w}) - \mathbf{z}_0)) - J^*\dot{J}\dot{\mathbf{w}} - (I_8 - J^*J)B_N\mathbf{w} - G_N(-B_N\mathbf{w} - K_N\mathbf{w})$$

- Consider the sequence of arrays w[n] at a rate h (e.g. 1msec), replace derivatives by finite differences, and after some algebra:

$$(I_8 + hA_1 + h^2A_2)\mathbf{w}[n] = \mathbf{w}[n-2] + 2\mathbf{w}[n-1] + hA_1\mathbf{w}[n-1] + h^2J^*K(\mathbf{z}_o - \mathbf{z}_c)$$

with: $$A_1 = -J^*BJ - J^*\dot{J} - B_N + J^*JB_N - G_NB_N$$
$$A_2 = -G_NK_N - J^*KF$$

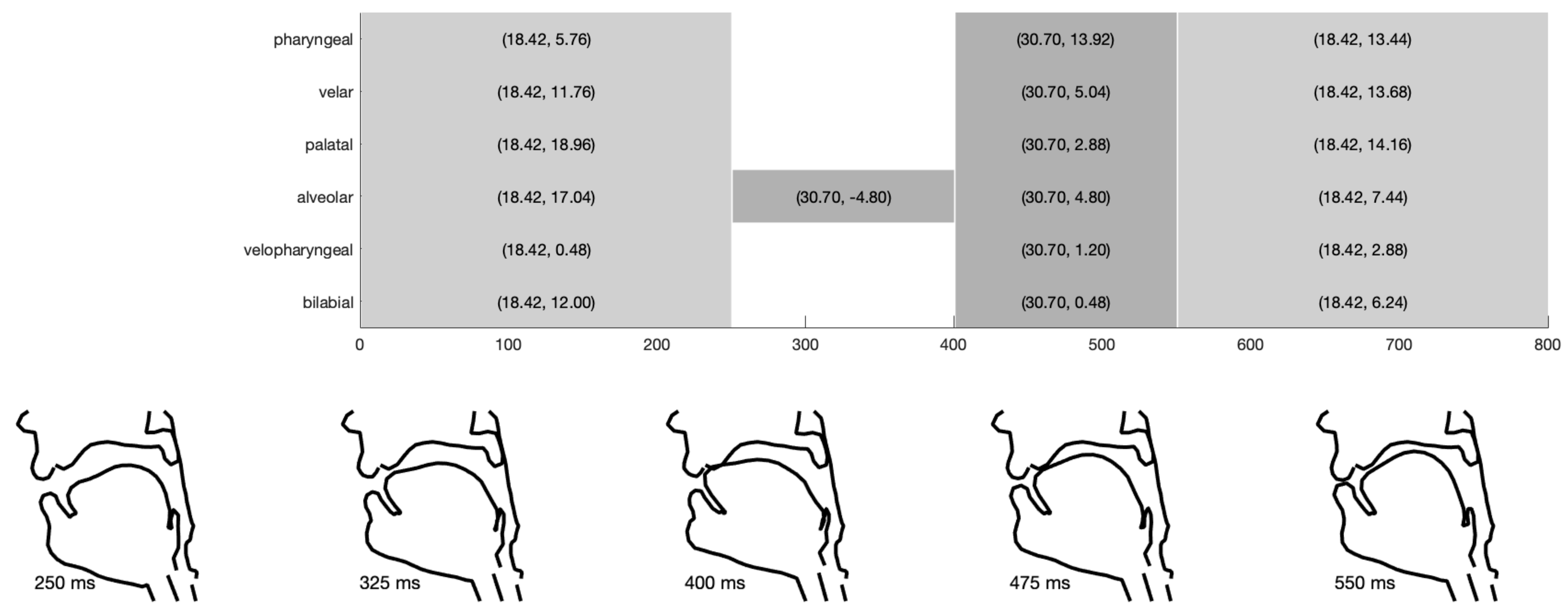
- Given a cluster we know F, J and its derivative, and we can invert (because the map in the cluster is linear) to get J*
- $z_o$ is a 6-dimensional array of **targets** and B, K are simple functions of a 6-dimensional array of natural frequencies $\omega_o$
- $G_N$ and $K_N$ are constants (neutral attractor)

## A Simple Loop

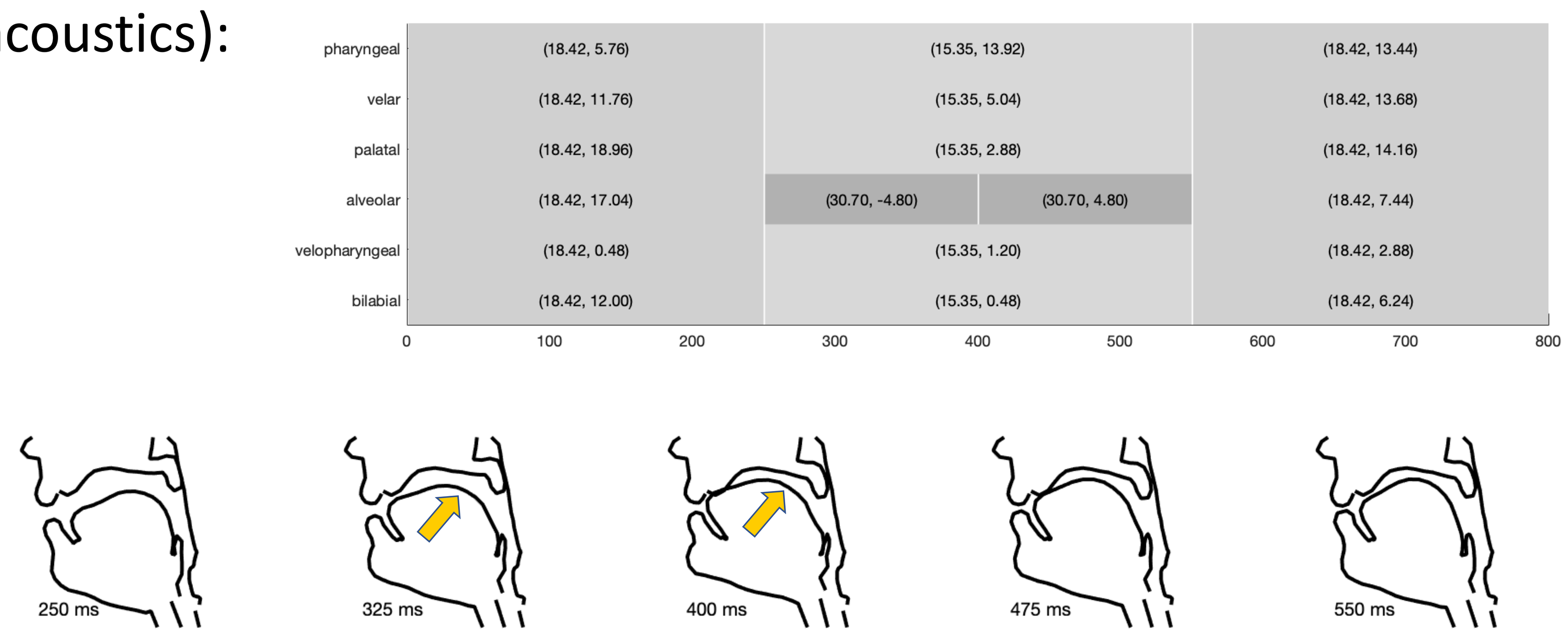
**Input:** $\mathbf{w}[0], \mathbf{w}[1], \boldsymbol{\omega}_o[n], \mathbf{z}_o[n], n = 2...N$
**for** $n$=2...$N$ **do**
  Find cluster where $\mathbf{w}[n-1]$ lies ;
  Retrieve $\mathbf{z_c}, F$ and $J$'s for that cluster ;
  Calculate $K(\boldsymbol{\omega}_o), B(\boldsymbol{\omega}_o)$;
  Solve system for $\mathbf{w}[n]$
**end**

## Simulating Anticipatory Coarticulation

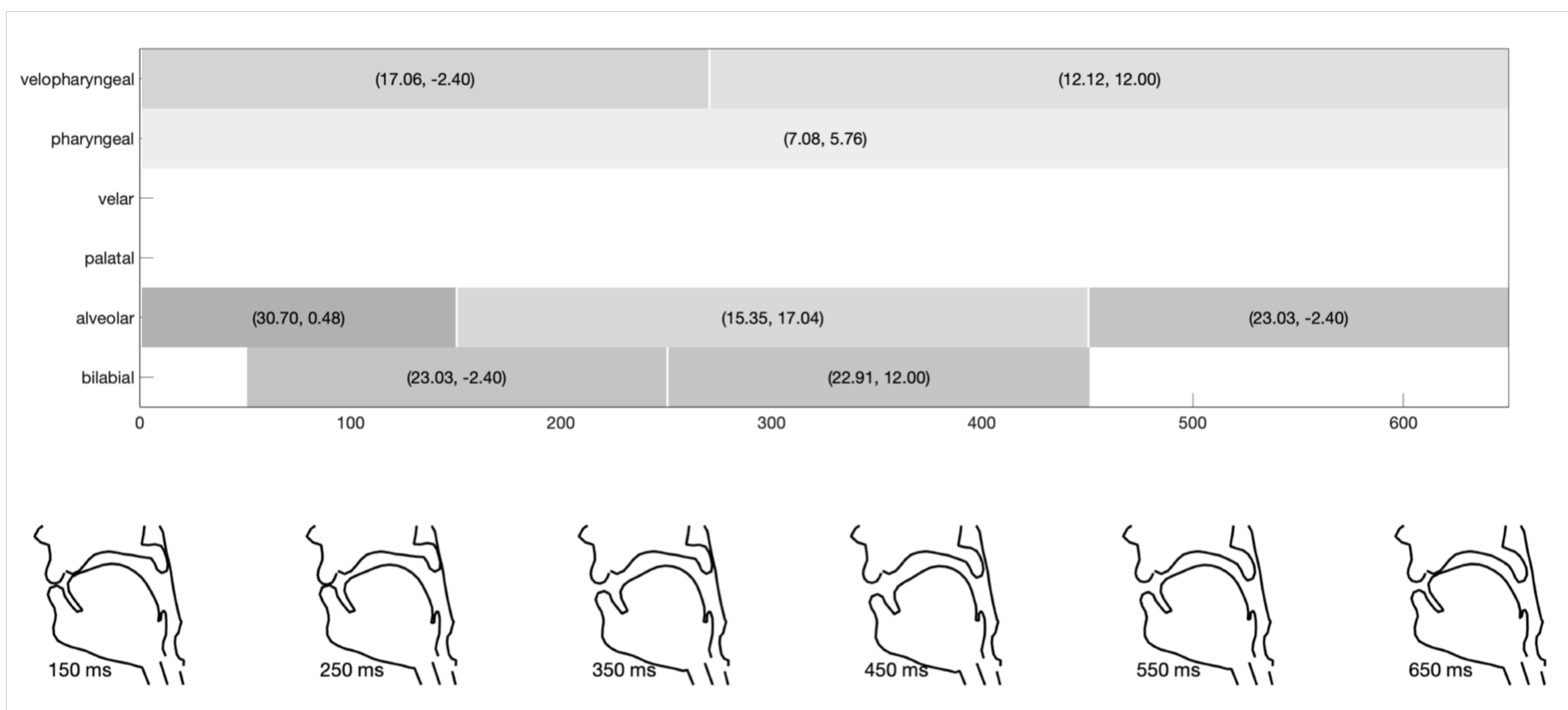
- Replicating Alexander et al. (2019), /adu/ without gestural overlap *(Note: In that work, targets for vowels were defined in parameter space, rather than constriction space)*




- Introducing anticipatory coarticulation, inspired by Ohman, leads to discernible difference in vocal-tract shaping dynamics (and also acoustics):




## Another Example: /span/




## Future Work

- Design gestural scores for more utterances
- Optimize scores to exactly fit recorded real-time MRI data