USC University of Southern California

USC Viterbi School of Engineering

Ming Hsieh Department of Electrical Engineering

Active or passive articulatory controls for emotion encoding?

Simulation of non-critical articulatory movements using critical movements in emotional speech

Jangwon Kim, jangwon@usc.edu, Electrical Engineering Dept. Signal Analysis and Interpretation Laboratory (SAIL)

Introduction

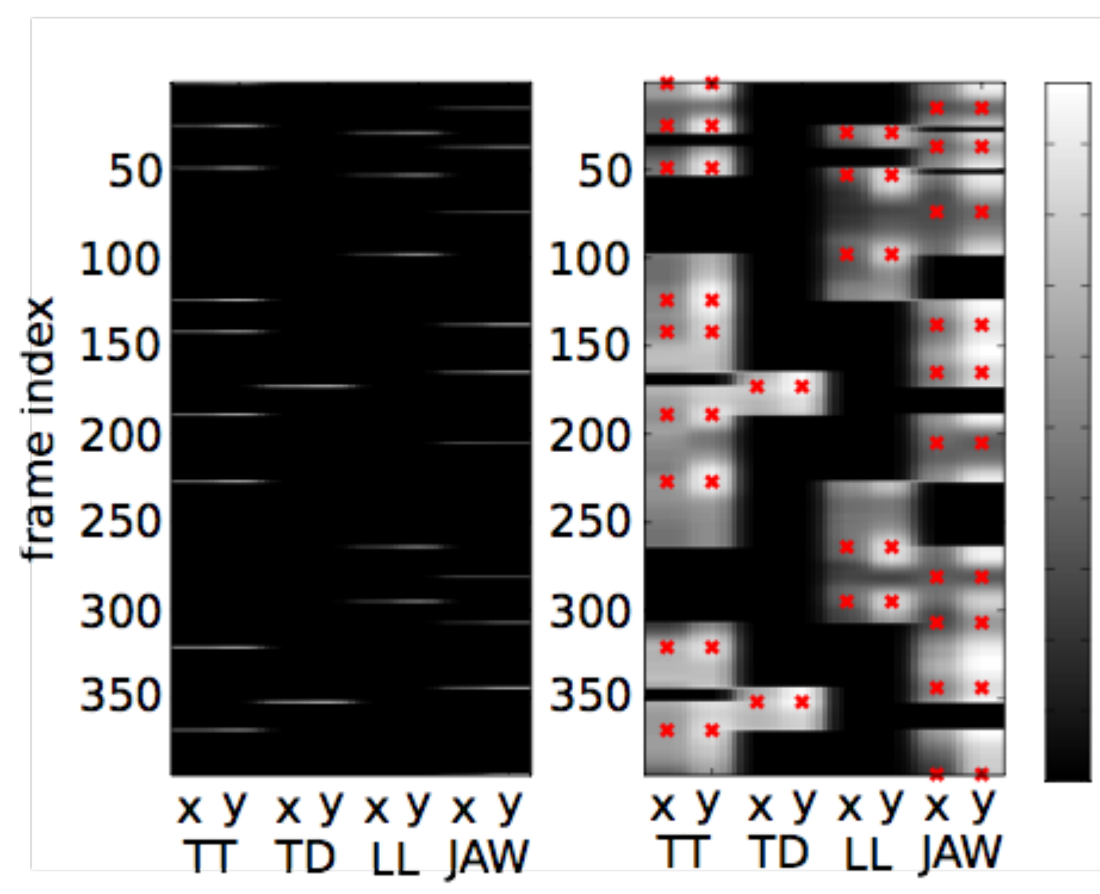
Background:

- In general, the movements of non-critical articulators display emotion variability more than the movements of critical articulators.
- For neutral speech, previous studies have assumed that the large variability of non-critical articulators' movements is the mechanical outcomes of the controls of critical articulators.

Objective:

- To answer the question whether the large emotional information displayed in the non-critical articulators are active control or passive control?
- to understand the underlying articulatory control mechanism in emotional speech, particularly for the emotional variation on the movements of non-critical articulators.

Critical / non-critical points



An example plot of critical time points (non-zero values) [left plot] and critical articulatory data (non-zero values) used for estimating

- Critical articulators: the articulators which are critial to produce a certain linguistic goal, i.e., phone
- Lips are more critical to produce /p/ than tongue: lips are critical articulators, and tongue is non-critical.

Non-critical movement model

$$\hat{f}_i(t) = f_i(t_c)K_i(t) + \hat{f}_i^p(t)(1 - K_i(t))$$

- $\hat{f}_i(t)$: Estimated position of i-th articulator at time t
- $K_i(t)$: weighting function on the nearest critical position
- $f_i(t_c)$: Contextual constraints of the nearest critical articulatory position
- $\hat{f}_i^p(t)$: Physiologically constrained motion (hidden)

t_c: nearest critical time point from t for the i-th articulator

$$K_i(t) = \frac{1}{1 + \exp(-\eta(\lambda_i(t) - \xi))}$$

$\lambda_i(t)$: time-varying influence from the nearest critical time point of the i-th articulator

η, ξ : hyper-parameters

$$\hat{f}_i^p(t) = \sum_{\substack{l=1 \\ l \neq i}}^{N_C(t)} (\alpha_{i,l} f_l(t)) + \beta_i$$

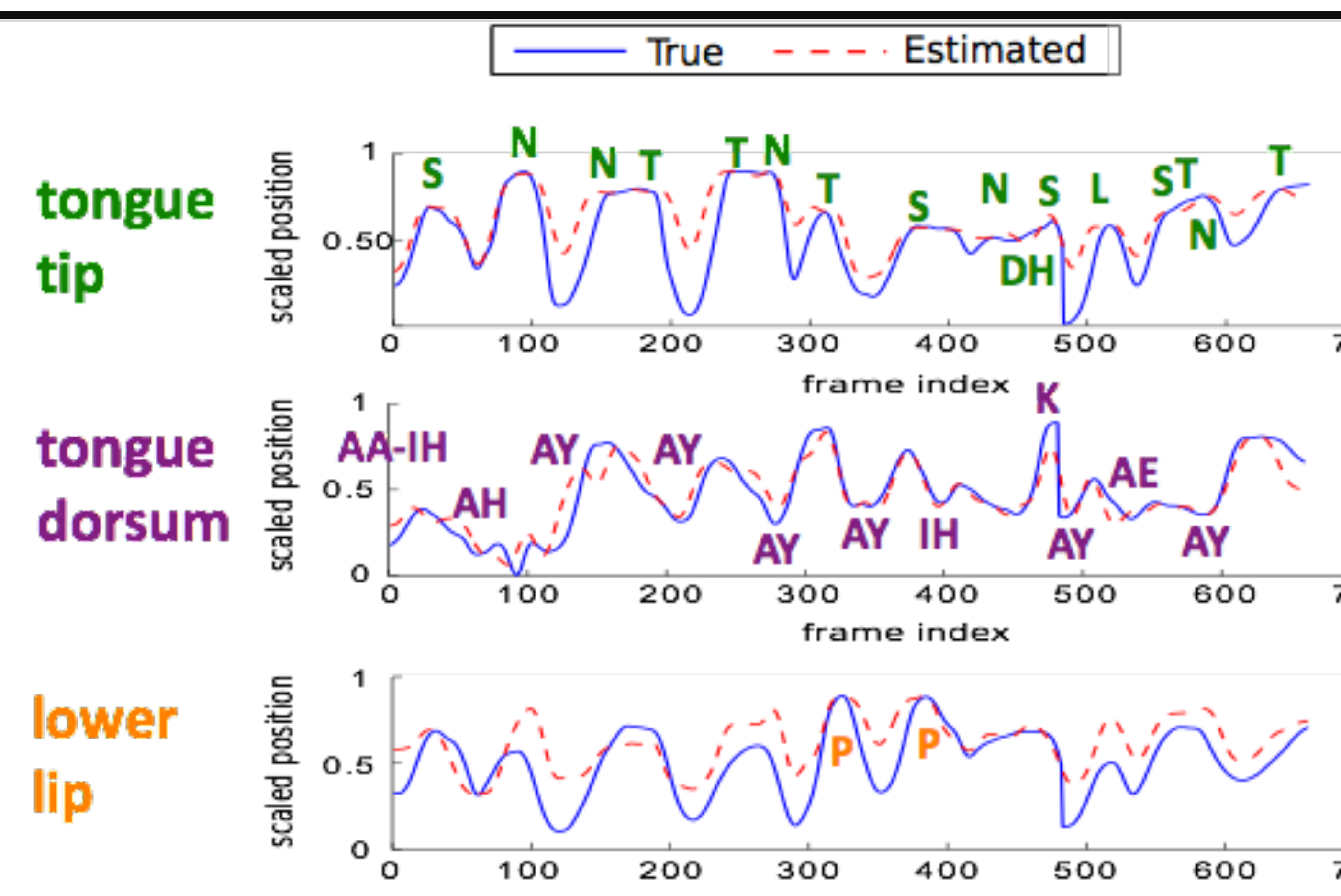
$f_l(t)$: position of l-th (currently critical) articulator

$N_C(t)$: The number of the critical articulators, except i-th articulator at t

$\alpha_{i,l}$: Coefficient reflects the dependency of the articulator i to the articulator j

β_i : constant error term

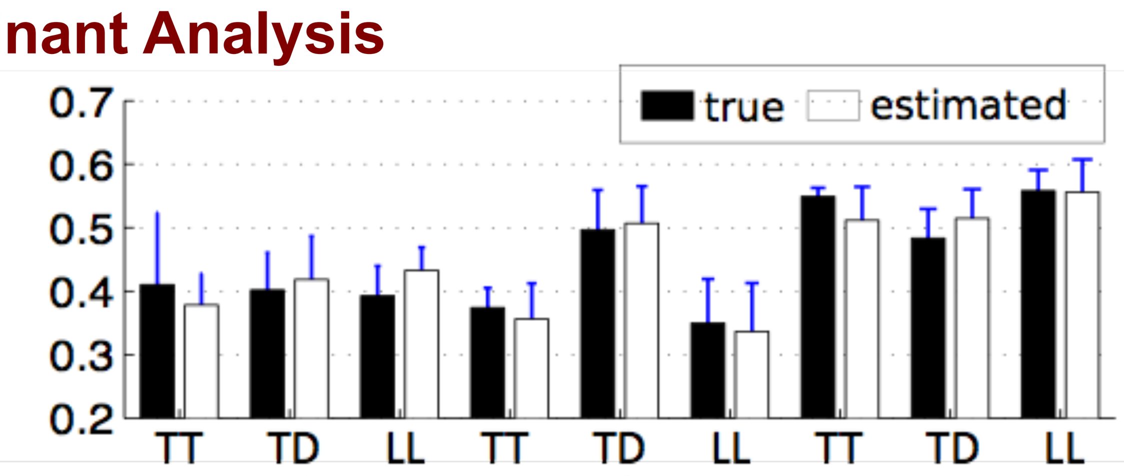
Experimental results



Between true and estimated non-critical trajectories…

- Mean (std) of RMSE: 0.069 (0.023) ~ 0.079 (0.030)
- Mean (std) of correlation coefficients: 0.794 (0.141) ~ 0.865 (0.112)

Discriminant Analysis



Emotion classification accuracy on true and estimated data.
2D normal density model, one mode for each emotion

⇒ **Similar classification accuracies on true and estimated data support the high dependency of emotional variation in non-critical articulatory movements to the control of critical articulatory movements.**

jangwon@usc.edu, funded by NSF

Ming Hsieh Institute
Ming Hsieh Department of Electrical Engineering