**Estimation of the movement trajectories of non-crucial articulators based on the detection of crucial moments and physiological constraints**

Jangwon Kim, Sungbok Lee, Shrikanth S. Narayanan
jangwon@usc.edu, sungbok@usc.edu, shri@sipi.usc.edu

[Supported by NSF IIS-1116076 and NIH DC007124]

**Introduction**

**Question:** Is the postural variability of non-crucial articulators a by-product of controls of crucial articulators and physiological constraints in the vocal tract or the results of active controls as a function of emotion?

**Objective 1:** Develop a mathematical model that estimates the movements of (linguistically) non-crucial articulators using the time and position of crucial articulators and physiological constraints among the articulators.
- For /p/, lips are crucial, and tongue points is non-crucial articulator.
- Eventually to understand the control mechanism of non-crucial articulators better.

**Objective 2:** Automatic detection of crucial articulatory moments on articulatory trajectories in an utterance.
- To reduce the time to determine time labels for phonemes on articulatory trajectories.
- Simple approach to relying on only physical properties of crucial articulators, because a large amount of phonetic labels on articulatory trajectories can not be easily obtained.

**Electromagnetic Articulography database**

- The NDI WAVE system
- 6 articulatory sensors: tongue tip (TT), tongue blade (TB), tongue dorsum (TD), upper lip (UL), lower lip (LL), lower incisor (JAW)
- 8 sentences x 5 repetitions
- A female native speaker of American English
- 5 acted emotions: neutrality, anger, happiness, sadness, fear
- Post processing for occlusal plane correction and smoothing on articulatory trajectories
- Emotion evaluation by 11 native speakers of American English

**Table 1:** The number of utterances for each emotion label

<table>
<thead>
<tr>
<th># Arti</th>
<th>Neutrality</th>
<th>Anger</th>
<th>Happiness</th>
<th>Sadness</th>
<th>Fear</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.09</td>
<td>0.10</td>
<td>0.09</td>
<td>0.08</td>
<td>0.11</td>
<td>200</td>
</tr>
<tr>
<td>6</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>0.78</td>
<td>0.76</td>
<td>0.77</td>
<td>0.80</td>
<td>0.80</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>0.87</td>
<td>0.84</td>
<td>0.85</td>
<td>0.85</td>
<td>0.85</td>
<td>42</td>
</tr>
</tbody>
</table>

**Figure 1:** O5(t) (red triangle line), and CSj(t) (blue asterisk line) and vertical trajectory (green square line) of crucial articulators for “nine one five.” Vertical dash-dot line (magenta) indicates the aligned crucial time point for each phone.

**Figure 2:** An example plot of Crucial time points (non-zero values) for each articulatory trajectory.

**Figure 3:** An example plot of the original and estimated vertical trajectories of the tongue tip.

**Forced alignment of crucial points**

- The proposed method uses constriction score CSj(t) and opening score O5(t) driven from the physical properties inherent in the articulatory movements. Crucality score Cj(t) is:

\[
C_j(t) = \frac{\theta_j(t) - \min(\theta_j)}{\max(\theta_j) - \min(\theta_j)} - \frac{S_j(t) - \min(S_j)}{\max(S_j) - \min(S_j)}
\]

where \( S_j = [S_j(1), S_j(2), \ldots, S_j(N)] \) is the sequence of the tangential speed of the i-th articulator; \( [\theta_1, \theta_2, \ldots, \theta_N] \) is the sequence of the acute angles of the i-th articulator; N is the number of frames for an utterance; t is the time frame.

- Finally, CSj(t) and O5j(t) are represented as a function of Cj(t) and a normalized local excursion score Ei(t):

\[
CS_j(t) = C_j(t) \times |E_i(t) - 1|
\]

\[
O5_j(t) = C_j(t) \times E_i(t)
\]

where \( E_i(t) \) is the degree of articulatory opening between two preceding and following extrema. \( E_i(t) \in [0, 1] \)

- Optimal crucial time points, one point for each phone, are determined by maximizing the sum of CSj(t) or O5j(t) (one score for each phone) using the Viterbi algorithm.

**Estimation of the trajectories of non-crucial articulators**

- \( f_i(t) \) and \( \hat{f}_i(t) \) are the true and estimated trajectory of i-th (non-crucial) articulator at time t.

\[
\hat{f}_i(t) = f_i(t) \cdot K_i(t) + \hat{f}_i(t)(1 - K_i(t))
\]

where \( K_i(t) \in [0, 1] \) is a weighting function on the contextual constrained motions.

- \( \hat{f}_i(t) \) is physiologically constrained motion of the i-th articulator. It is modeled by a linear regression of the positions of all crucial articulators at time t. Ex: Linear transformation of positions of the jaw and the lower lip.