Motivation & Introduction

Motivation: Difficulty of articulatory data collection of emotional speech, the (potential) benefits of the estimated articulatory data (complementary information gain to other emotional features from acoustic signal)

Goal: Understanding the effectiveness of acoustic-to-articulatory inversion technique for emotional speech.

Research Questions (RQ):
(1) How similar is the estimated trajectory to the original articulatory trajectory for various emotions?
(2) How much does the estimated trajectory preserve emotion specific information (ESI)?

Results and Discussion

Average MSE of two trajectories:

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Original</th>
<th>Estimated</th>
<th>aveDiff</th>
<th>F</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neu</td>
<td>2.26</td>
<td>3.82</td>
<td>-1.56</td>
<td>335</td>
<td>0.000</td>
</tr>
<tr>
<td>HAn</td>
<td>3.88</td>
<td>3.92</td>
<td>-0.04</td>
<td>10</td>
<td>0.001</td>
</tr>
<tr>
<td>Cang</td>
<td>2.81</td>
<td>3.95</td>
<td>-1.14</td>
<td>70</td>
<td>0.000</td>
</tr>
<tr>
<td>Hap</td>
<td>3.52</td>
<td>3.88</td>
<td>-0.36</td>
<td>35</td>
<td>0.000</td>
</tr>
<tr>
<td>Sad</td>
<td>2.91</td>
<td>3.85</td>
<td>-0.94</td>
<td>35</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Experimental Setup

For RQ1: comparison of two articulatory trajectories (original-estimated) using the MSE measure of the same emotion pair with that of different emotion pair.

For RQ2: to test using emotion classification accuracy.

Generalized smoothness criterion based inversion
-Acoustic-to-articulatory inversion used here
-Constraining individual articulatory trajectories by the corresponding articulator-specific smoothness factors.

Database: Electromagnetic articulography
-Parallel waveform and articulatory trajectory
-5 categorical & elicited emotions
-2 females, 1 male spoke 7 sent, 3 styles, 4–5 rep.
-6 articulatory points (3 for tongue, 2 for lips, 1 for jaw)

Conclusions & Future Work

Conclusions:
-Articulatory data estimated using GSC based inversion carry important ESI, but smaller than original.
-ESI in the estimated trajectories offers complementary emotion information to that in the acoustic prosodic features.
-Remaining Question: how ESI is encoded in the estimated articulatory movements (in dimensional or categorical way of emotion?)

Future work directions:
-Investigate other inversion techniques
-Subject-independent inversion for emotional speech

Emotion Classification:

GMM

SVM

-Estimated articulatory data has emotion-specific information (ESI) but not to the same degree as in orig. articulatory data.
-Estimated articulatory data has ESI complementary to the acoustic and prosodic features.

Contact information, sponsors etc.