# USC-EMO-MRI corpus: An emotional speech production dataset recorded by real-time magnetic resonance imaging

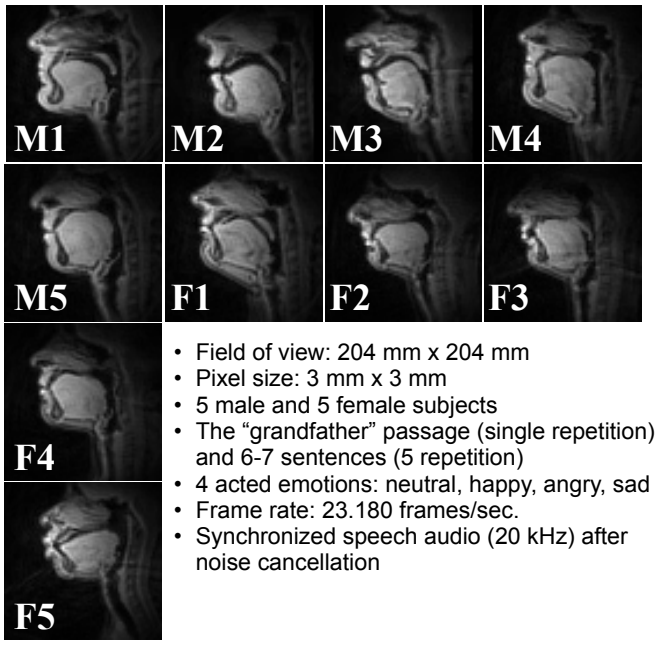
*Jangwon Kim, Asterios Toutios, Yoon-Chul Kim,*
*Yinghua Zhu, Sungbok Lee and Shrikanth S. Narayanan*
*Signal Analysis and Interpretation Lab, University of Southern California*
*This work is supported by NSF IIS-1116076 and NIH DC007124*

## Motivation

- Providing a resource for systematic analysis for the ***inter- and intra-speaker variability*** of emotional speech in the articulatory movements and prosodic behaviors.
- Assisting ***a comprehensive modeling*** of the vocal tract shaping, and eventually the ***joint modeling*** of articulatory and acoustic behaviors with emotion coloring.

## Data collection

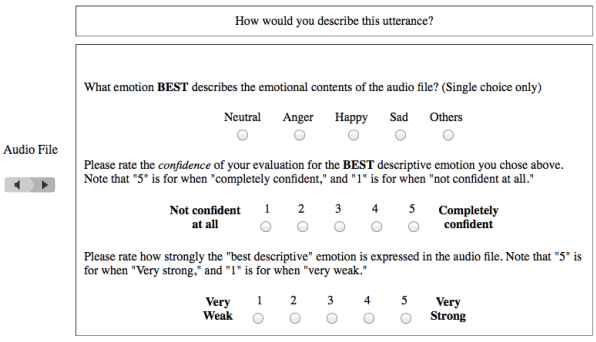
### •Real-time Magnetic Resonance Imaging (rtMRI)




- Field of view: 204 mm x 204 mm
- Pixel size: 3 mm x 3 mm
- 5 male and 5 female subjects
- The “grandfather” passage (single repetition) and 6-7 sentences (5 repetition)
- 4 acted emotions: neutral, happy, angry, sad
- Frame rate: 23.180 frames/sec.
- Synchronized speech audio (20 kHz) after noise cancellation

### •Emotion evaluation

(e.g., http://sail.usc.edu/~jangwon/mri_nsf_eval_jr_short)



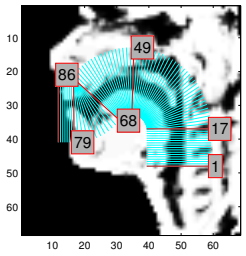

| | | Perceived | | | | |
|---|---|---|---|---|---|---|
| | | Neutrality | Anger | Happiness | Sadness | Other |
| Target | Neutrality | 58 | 0 | 0 | 0 | 4 |
| | Anger | 2 | 42 | 1 | 0 | 7 |
| | Happiness | 1 | 0 | 38 | 0 | 13 |
| | Sadness | 2 | 0 | 0 | 39 | 11 |
| | Total | 63 | 42 | 39 | 39 | 35 |

| | Subject ID (M: male, F: female) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | M1 | M2 | M3 | M4 | M5 | F1 | F2 | F3 | F4 | F5 |
| #Eval | 10 | 10 | 11 | 10 | 11 | 12 | 12 | 12 | 12 | 10 |
| Sent | 1-6 | 1-7 | 1-7 | 1-7 | 1-7 | 1-6 | 1-6 | 1-6 | 1-7 | 1-7 |
| AVE | 85.3 | 69.5 | 82.6 | 72.0 | 80.5 | 80.7 | 94.0 | 89.8 | 86.5 | 80.5 |
| STD | 9.9 | 11.8 | 8.5 | 11.1 | 10.0 | 11.1 | 5.4 | 8.4 | 8.2 | 11.8 |

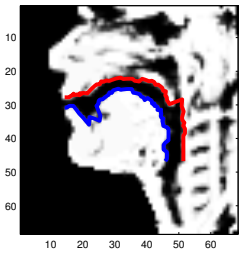
## Preliminary analysis of articulatory variation depending on emotion

### • Automatic MR image segmentation

*http://sail.usc.edu/old/software/rtmri_seg*



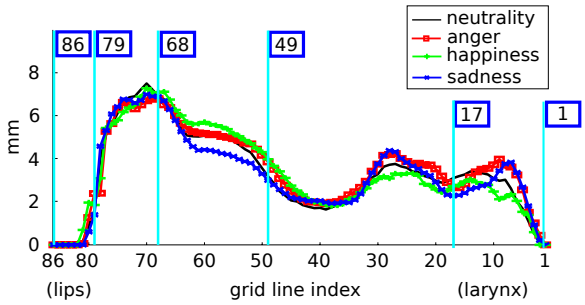

*Grid line based analysis*



*Airway-tissue boundaries*

- Semi-automatically determined grid lines for analyzing the vocal tract shape
- Automatically segmented airway-tissue boundaries based on the grid lines
- Computing distance function, i.e. the Euclidean distance between the upper and lower boundaries for each grid line

### •Mean distance function for different emotions




*Mean of the range of distance functions for “FIVE”*

- **High v.s low arousal emotions:** Anger and happiness show wider movement range than sadness in the hard palate region (49-68) **=> Wider palatal opening for high arousal emotions is well captured.**
- **High v.s. low valence emotions:** Anger shows less movement range than happiness in the pharyngeal region (5-20), while anger shows more movement range than happiness in the hard palate region (49-68). **=> The pharyngeal constriction and releasing were emphasized for anger, while palatal constriction and releasing were more emphasized for happiness.**

## Conclusion and Future Works

- Variation of articulatory movement range depending on emotion is observed in this dataset.
- ”Lab” speech, “acted” emotion, MRI scanning environment.
- To collect parallel EMA data (same subject and stimuli).
- To add emotion evaluation from naive listeners.