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)

Preliminary analysis of articulatory variation depending on emotion
- Automatic MR image segmentation
  - 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

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.