1. Summary

- We hypothesize that certain speaker gestures can convey significant information that are correlated to audience engagement.
- Gesture attributes: velocity, direction and extremal pose.
- Measured correlation between gesture attributes and engagement: statistically significant correlations found, with Spearman correlation as high as 0.283 with \( p < 0.5 \), Bonferroni corrected.

2. Experiment: Visual Stimuli and Subjects

- Visual stimuli: Subjects shown 61 minutes of video clips in an RF-shielded room wearing 64-electrode EEG scalp cap.
  - 47 minutes of EEG recordings retained after discarding noisy data.
  - Clips created from U.S. Presidential debates semi-automatically:
    - Each clip featured a single speaker (Obama or Romney).
  - 28 clips selected and verified versions created.
  - Silent and audible versions combined into 6 longer videos each approximately 10 minutes long.

- Subjects: balanced by age, gender, political affiliation.
  - 5 of each of (male, female) X (Democrat, Republican).
  - Subjects surveyed for but not selected by topic interest.
  - Democrats and Republicans have similar levels of interest in topics
  - Males more indifferent than females.

3. Gesture Attributes

- 1. Velocity: How far each hand moves between frames (pixels).
- 2. Direction change: PCA on orientation of motion vectors between window of frames.
  - Angular distance between main components across frames.
  - E.g., “jitter”: high direction change, low velocity, or “swipe”: low direction change, high velocity.
- 3. Extremal pose: Track hand positions across full debate videos (stratify by speaker and debate)
  - Train GMMs to model position of hands
  - Number of centers set to number of hands
  - Identifies when hands tend to be in unusual positions

4. Audience Engagement from EEG

- Identifies time segments of maximum correlation between subjects’ neural activity [Dmochowski, et al. 2012]
- Let \( X_i \in \mathbb{R}^{N \times D} \) be the EEG data of subject \( i \) where \( D \) is the number of channels and \( T \) is the number of time samples
- For \( N \) subjects, form aggregated matrices:
  \[
  X_1 = \begin{bmatrix}
  X^{1}(1) & \cdots & X^{1}(N) & \cdots & X^{1}(N-2) & \cdots & X^{1}(N-3) & \cdots & X^{1}(N-1) \\
  \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\
  X^{N}(1) & \cdots & X^{N}(N) & \cdots & X^{N}(N-2) & \cdots & X^{N}(N-3) & \cdots & X^{N}(N-1) \\
  \end{bmatrix} \\
  X_2 = \begin{bmatrix}
  X^{1}(1) & \cdots & X^{1}(N) & \cdots & X^{1}(N-2) & \cdots & X^{1}(N-3) & \cdots & X^{1}(N-1) \\
  \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\
  X^{N}(1) & \cdots & X^{N}(N) & \cdots & X^{N}(N-2) & \cdots & X^{N}(N-3) & \cdots & X^{N}(N-1) \\
  \end{bmatrix}
  \] 

- We wish to find weight vector \( w \in \mathbb{R}^{N} \) such that Pearson correlation between \( Y_i = X_i w \) and \( Y_i = X_i w \) is maximized.
- Compute Pearson correlation between 5-second overlapping windows (at 1-second intervals) for weight vector
- Correlations computed from three weight vectors corresponding to strongest correlations referred to as components 1, 2, 3

5. Correlations

- Computed correlations between 3 gesture attributes (12 Hz) and 30 engagement features (1 Hz upscaled to 12 Hz).
- (5 subject groups) \( X \) (silent/audible) \( X \) (3 EEG components)
- Spearman correlation \( \rho \) used instead of Pearson.
- No reason extreme engagement should correlate with extreme gesture.
- Statistically significant correlations for all subjects shown.
- Permutation test; Bonferroni corrected.
- More subject stratifications in paper.

6. Observations

- Gestures are significant and may augment speech.
- Statistically significant correlations between gesture and engagement exist during both silent and audible playback.
- Extremal poses and direction change are the most significant gesture attributes.
- First and second components of engagement features drive correlation.
- Audible videos are more engaging in general, but silent videos are more likely to engage via gesture.
- Obama did not engage audience through gesture in first debate, whereas Romney did so in both debates.
- Consistent with media reports that Obama “lost” the first debate.

---

John R. Zhang\textsuperscript{1}, Jason Sherwin\textsuperscript{1}, Jacek Dmochowski\textsuperscript{2}, Paul Sajda\textsuperscript{1} and John R. Kender\textsuperscript{1}
\textsuperscript{1}Columbia University, New York NY \textsuperscript{2}Stanford University, Stanford CA