Gender Representation in Cinematic Content: A Multimodal Approach

Tanaya Guha, Che-Wei Huang, Naveen Kumar,
Yan Zhu, Shrikanth S. Narayanan

Signal Analysis and Interpretation Lab (SAIL)
University of Southern California, Los Angeles

- Women are underrepresented in popular movies and media

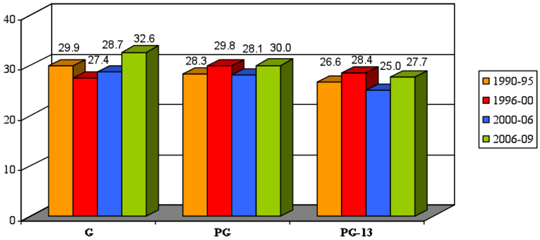

Figure: Female characters (%) in Hollywood movies [Smith et al. 2012]

- Women have fewer dialogs, and stereotypical appearance
- Watching biased content has negative impact on society

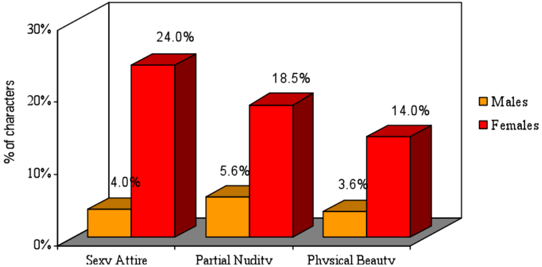

Figure: Comparison of physical appearance [Smith et al. 2012]

- Social scientists analyze movies *manually*.

- Social scientists analyze movies *manually*.
- Automated analysis can discover hidden trends and biases.

- Social scientists analyze movies *manually*.
- Automated analysis can discover hidden trends and biases.
- Can analyze big data

- Social scientists analyze movies *manually*.
- Automated analysis can discover hidden trends and biases.
- Can analyze big data
- **Goals:**
    - Objectively understand gender representation
    - Develop assistive tools for media informatics

- How to quantify gender representation/bias? No simple answer.
- Begin with simple measurable quantities
  - Screen presence from video
  - Speaking time from audio
- Leverage existing computer vision and speech processing techniques
- Combine and analyse multimodal cues

On-screen time estimation as a shot labeling problem

(Assumption: Characters do not appear or leave within a shot)

1. **Shot boundary detection**

On-screen time estimation as a shot labeling problem

(Assumption: Characters do not appear or leave within a shot)

1. **Shot boundary detection**

2. **Face detection**

- Knowledge-driven grouping of the tasks

  emotion groups: *anger, disgust, fear, joy, sadness and surprise*

- Knowledge-driven grouping of the tasks

  emotion groups: *anger, disgust, fear, joy, sadness and surprise*

- Differences between ASD and TD for each emotion group
  - Global analysis: Dynamical complexity

    multivariate multiscale entropy (MMSE)

  - Local analysis: region-based dynamics

    psychology-inspired measures

    time series modeling

- Presence in key parts of a movie counts more
- Segment movie into scenes
- Estimate 'importance' of each scene from multimodal cues
  - video stream: *average shot length*, *motion activity*
  - audio stream: *harmonicity*

- 17 full-length Hollywood movies
- Independent analysis of video and audio
- $B^m(r)$ = fraction of the composite vector pairs for which $d(C_i, C_j) \leq r$, where $i \neq j$
- $B^{m+1}(r)$ = same measure for (m+1)-dim space
- Includes vector pairs within and across the embedded subspaces

- Divide markers in 8 facial regions:
  Left and Right Eye brow, Eye, Cheek and Mouth

- Divide markers in 8 facial regions:
  Left and Right Eye brow, Eye, Cheek and Mouth
- Compute psychology-inspired measures
  - mimicry quality
  - left-right (motion) activation symmetry
  - upper-lower (motion) activation divergence

- Stimuli:
  - Experts score each region (0 - 5) depending on how activated is a region during an expression
  - Human annotated motion activation vector: $\mathbf{v} = [v_1, v_2, ..., v_8]$

- Stimuli:
  - Experts score each region (0 - 5) depending on how activated is a region during an expression
  - Human annotated motion activation vector: $\mathbf{v} = [v_1, v_2, ..., v_8]$
- Subjects:
  - Motion activation measured from participant's mocap data
  - Motion activation$_{reg}$ = total distance traveled by all markers from their respective rest positions
  - Let $\mathbf{a} = [a_1, a_2, ..., a_8]$ be the activation vector

- Autoregressive (AR) model of order p to model the dynamics of each region:

$$d_t = \sum_{i=1}^{p} \alpha_i d_{t-i} + \sigma_t, \tag{1}$$

  where σ_t is white noise, $\{\alpha_i\}_{i=1}^{p}$ are the model parameters
- model order $= 4$

- Contribution
  - Quantification of facial-expression related atypicality in autism
  - Identifying expressions which induce more awkwardness
  - Identifying the facial regions inducing perception of atypicality

SUMMARY

- Contribution
  - Quantification of facial-expression related atypicality in autism
  - Identifying expressions which induce more awkwardness
  - Identifying the facial regions inducing perception of atypicality
- Observations
  - ASD group has reduced dynamical complexity supporting the well-known complexity loss theory
  - Differences in dynamics is higher in eye region
  - ASD subjects underperform in mimicry
  - Differences are pronounced for expressions of disgust and sadness (negative valence)

Thank you

NSF

Geena Davis Institute on Gender in Media

Google