First Grand-Challenge and Workshop on Human Multimodal Language (Challenge-HML)

Organizers:

Amir Zadeh | Louis-Philippe Morency | Paul Pu Liang | Soujanya Poria | Erik Cambria | Stefan Scherer | Zhun Liu
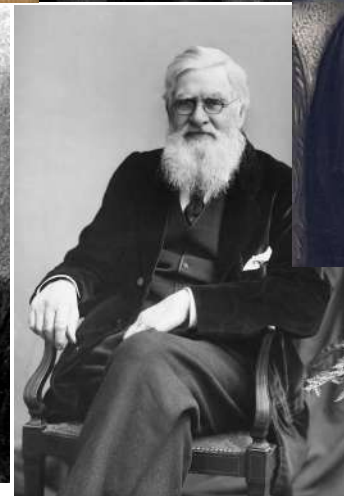
Continuous Theories of (Multimodal) Language
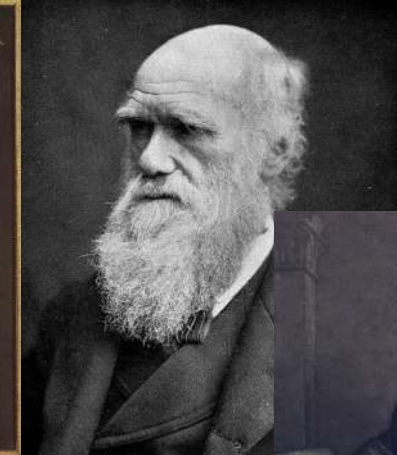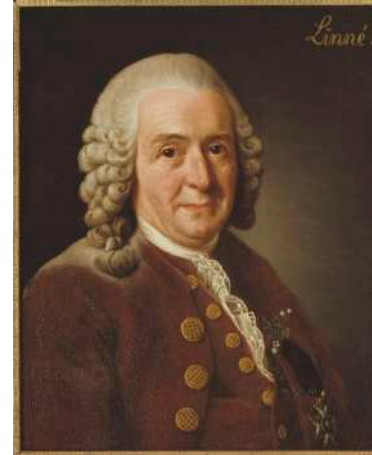
Throughout evolution language and nonverbal behaviors developed together.

Cries and Imitations   Modern Language
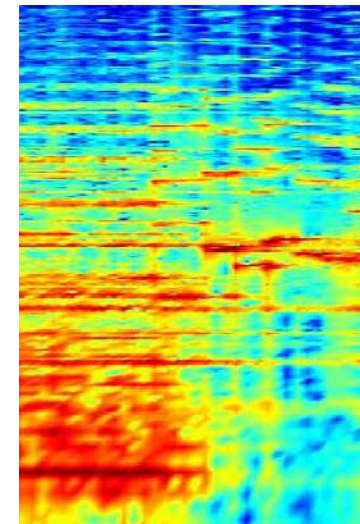
Multimodal Language Modalities

"I really like this vacuum cleaner, it was one of the few that has a large dust bag ..."

Language (words)

Vision (gestures)

Acoustic (voice)

Challenge-HML Structure

▪ Our goal in the first Challenge-HML is to build models of sentiment and emotions for human multimodal language through the proxy of in-the-wild speech videos.
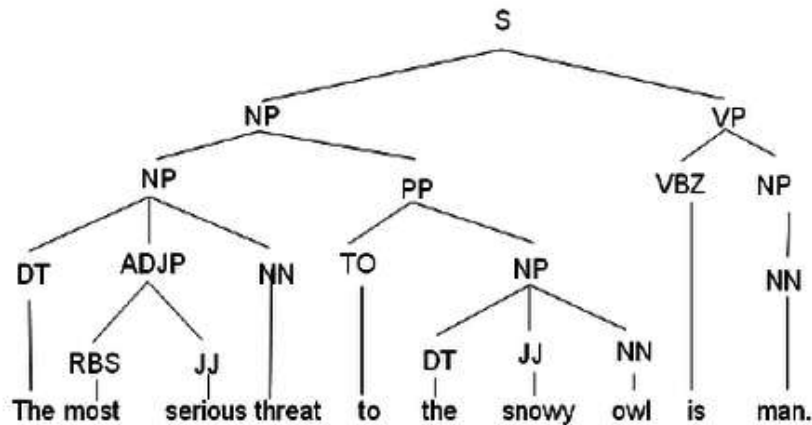
Challenge-HML Metrics

▪ Sentiment: Binary and Multiclass Sentiment Analysis

▪ Emotion Recognition: Binary and Multiclass Analysis of
- Happiness
- Sadness
- Anger
- Surprise
- Disgust
- Fear

Difficulties in Modeling Multimodal Language

- Modeling
  - Intra-modal Dynamics: difficulties in modeling each modality.
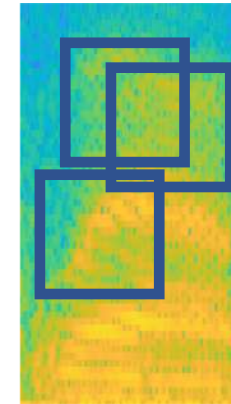
Language (words)   Vision (gestures)   Acoustic (voice)

Difficulties in Modeling Multimodal Language

- Modeling
  - Intra-modal Dynamics
    - Difficulties in modeling each modality
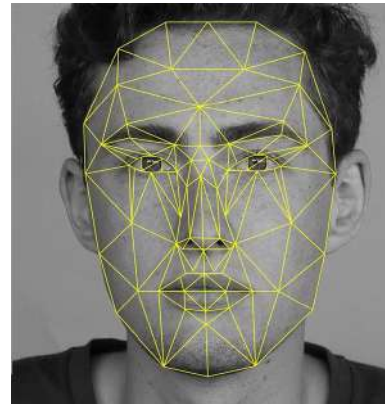 - Inter-modal Dynamics (fusion): difficulties in modeling spatio-temporal relations between modalities
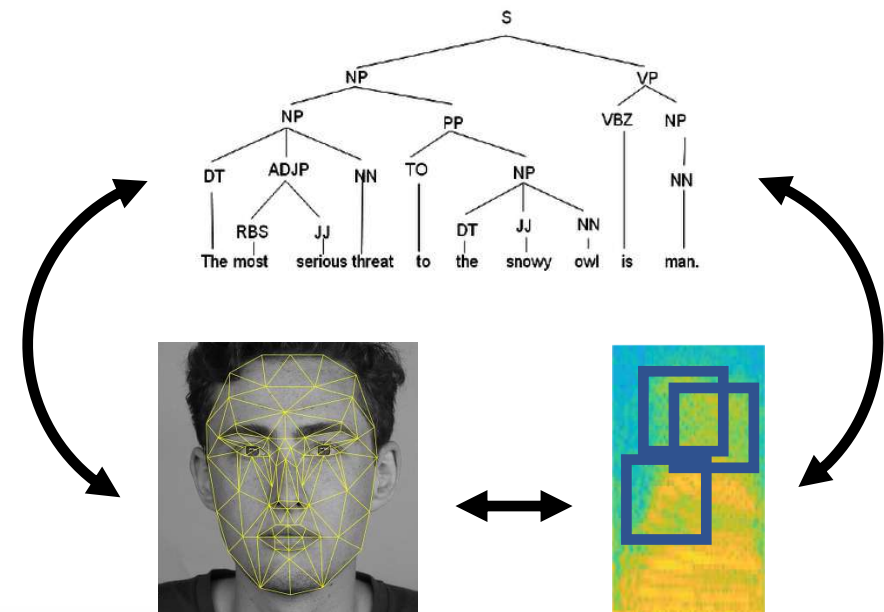
Difficulties in Modeling Multimodal Language

- Modeling
  - Intra-modal Dynamics
    - Difficulties in modeling each modality
  - Inter-modal Dynamics (fusion)

- Idiosyncratic signal
  - Speakers talk and behave differently
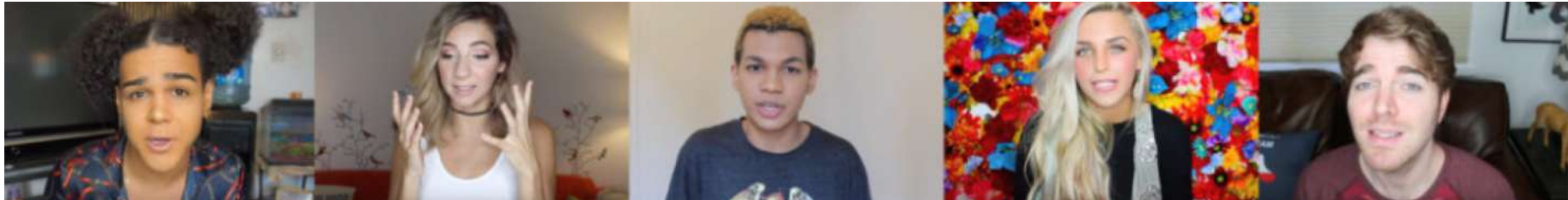
CMU-MOSEI Dataset

- CMU-Multimodal Sentiment and Emotion Intensity Analysis Dataset

  - Largest multimodal dataset of sentiment and emotion analysis.
  - More than 23,000 sentence utterances from YouTube.
  - More than 3,000 YouTube videos.
  - More than 1000 speakers.
  - More than 250 topics.
  - All three modalities.
  - Manual transcriptions.
  - Phoneme-level Alignment (semi-automatic)
  - Annotated for sentiment and emotions (Likert scales)
    - 7-scale sentiment (CMU-MOSI, SST)
    - 4-scale emotions (happiness, sadness, anger, disgust, surprise, fear)

CMU-MOSEI Dataset

Dataset	# S	# Sp	Mod	Sent	Emo	TL (hh:mm:ss)
CMU-MOSEI	**23,453**	**1,000**	$\{l,v,a\}$	✓	✓	**65:53:36**
CMU-MOSI	2,199	98	$\{l,v,a\}$	✓	✗	02:36:17
ICT-MMMO	340	200	$\{l,v,a\}$	✓	✗	13:58:29
YouTube	300	50	$\{l,v,a\}$	✓	✗	00:29:41
MOUD	400	101	$\{l,v,a\}$	✓	✗	00:59:00
SST	11,855	–	$\{l\}$	✓	✗	–
Cornell	2,000	–	$\{l\}$	✓	✗	–
Large Movie	25,000	–	$\{l\}$	✓	✗	–
STS	5,513	–	$\{l\}$	✓	✗	–
IEMOCAP	10,000	10	$\{l,v,a\}$	✗	✓	11:28:12
SAL	23	4	$\{v,a\}$	✗	✓	11:00:00
VAM	499	20	$\{v,a\}$	✗	✓	12:00:00
VAM-faces	1,867	20	$\{v\}$	✗	✓	–
HUMAINE	50	4	$\{v,a\}$	✗	✓	04:11:00
RECOLA	46	46	$\{v,a\}$	✗	✓	03:50:00
SEWA	538	408	$\{v,a\}$	✗	✓	04:39:00
SEMAINE	80	20	$\{v,a\}$	✗	✓	06:30:00
AFEW	1,645	330	$\{v,a\}$	✗	✓	02:28:03
AM-FED	242	242	$\{v\}$	✗	✓	03:20:25
Mimicry	48	48	$\{v,a\}$	✗	✓	11:00:00
AFEW-VA	600	240	$\{v,a\}$	✗	✓	00:40:00

Keynotes

Dr. Bing Liu
University of Illinois at Chicago (USA)

Dr. Sharon Oviatt
Monash University (Australia)

Dr. Roland Göcke
University of Canberra (Australia)

Thank you for joining us today