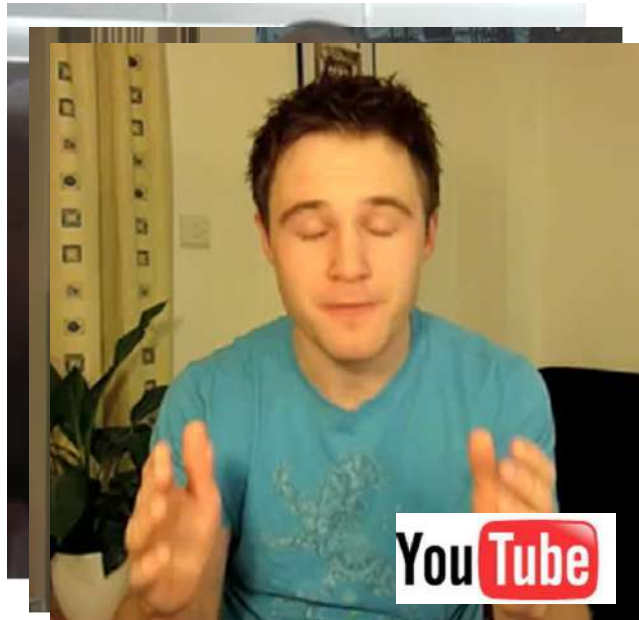# Computational Modeling of Human Multimodal Language:
The MOSEI Dataset and Interpretable Dyanamic Fusion
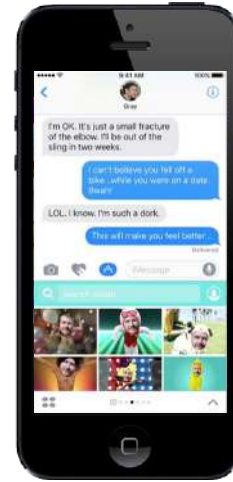
Paul Pu Liang

Committee: Louis-Philippe Morency, Ruslan Salakhutdinov

Progress of Artificial Intelligence

Multimedia Content

Intelligent
Personal Assistants

Robots and
Virtual Agents

Multimodal Communicative Behaviors

Language
- **Lexicon**
- **Syntax**
- **Pragmatics**

Acoustic
- **Prosody**
- **Vocal expressions**

Visual
- **Gestures**
- **Body language**
- **Eye contact**
- **Facial expressions**

Sentiment
- Positive
- Negative

Emotion
- Anger
- Disgust
- Fear
- Happiness
- Sadness
- Surprise
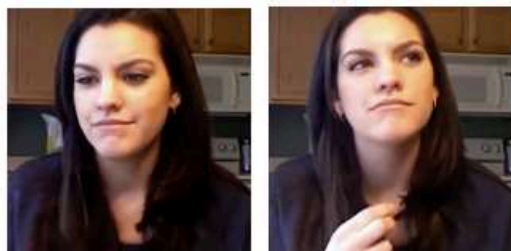
Personality
- Confidence
- Persuasion
- Passion

New Dataset: MOSEI

23,000 video segments
3 modalities

Language: And he I don't think he got mad when hah I don't know maybe. | Too much too fast, I mean we basically just get introduced to this character... | All I can say is he's a pretty average guy.
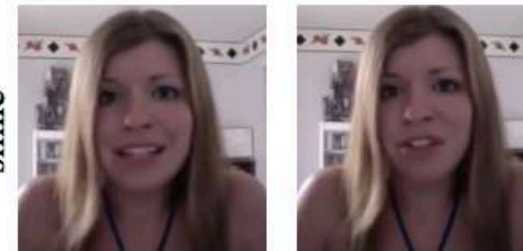
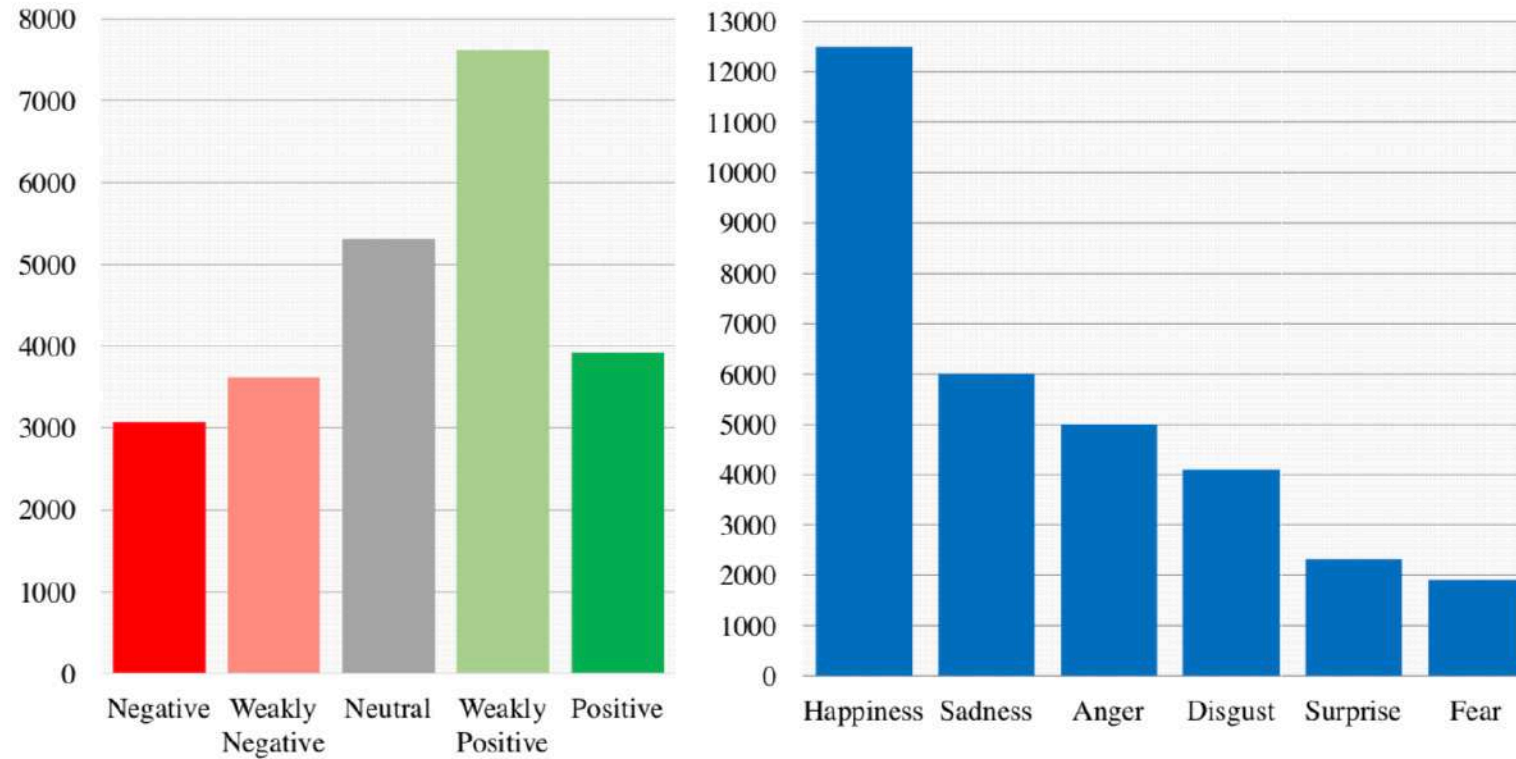Vision: Gaze aversion | Uninformative | Contradictory smile
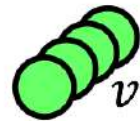
Acoustic: (frustrated voice) | (angry voice) | (disappointed voice)

MOSEI Dataset

1,000 speakers   250 topics

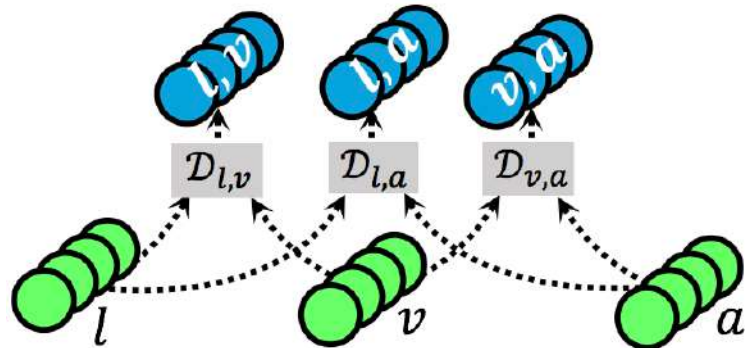# Annotation Distributions

Dynamic Fusion Graph
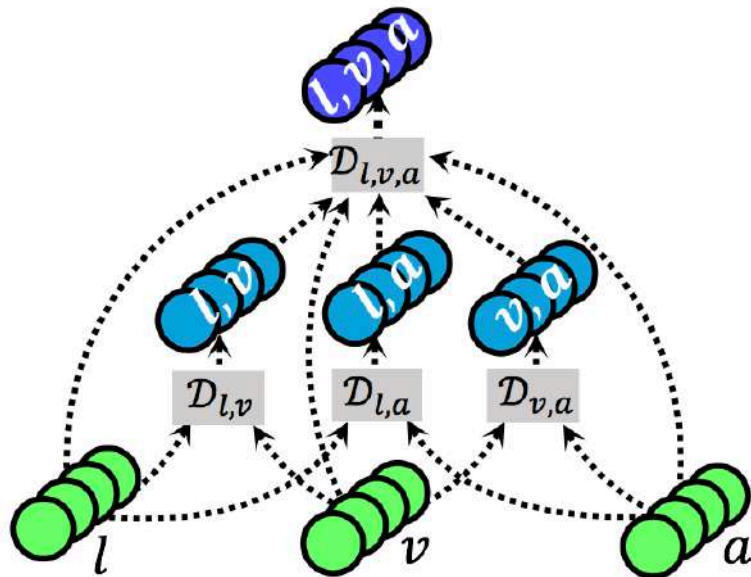

unimodal

Dynamic Fusion Graph

bimodal

unimodal

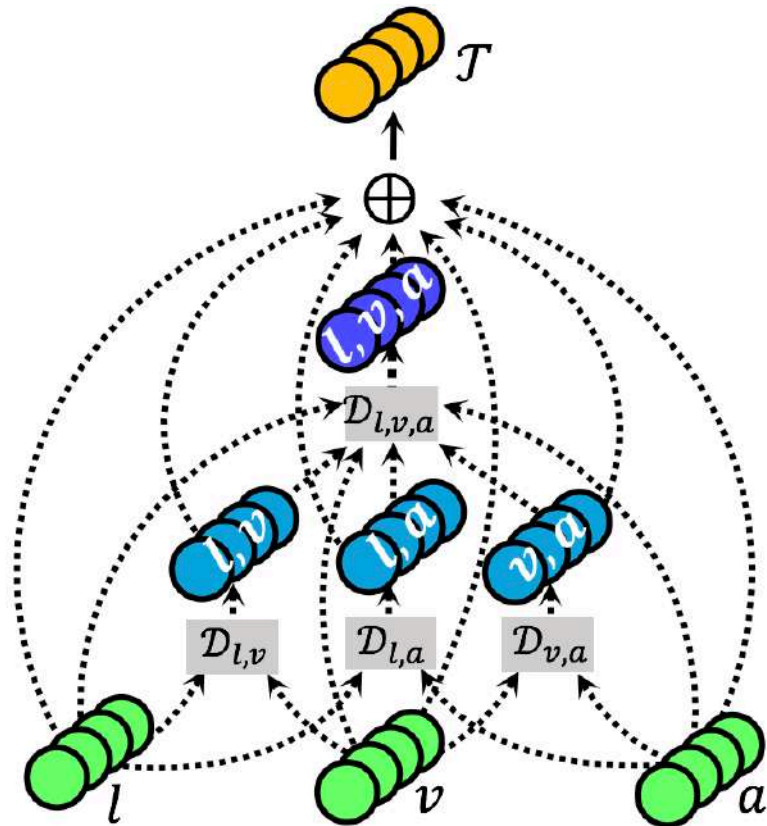# Dynamic Fusion Graph

trimodal

bimodal

unimodal

Dynamic Fusion Graph
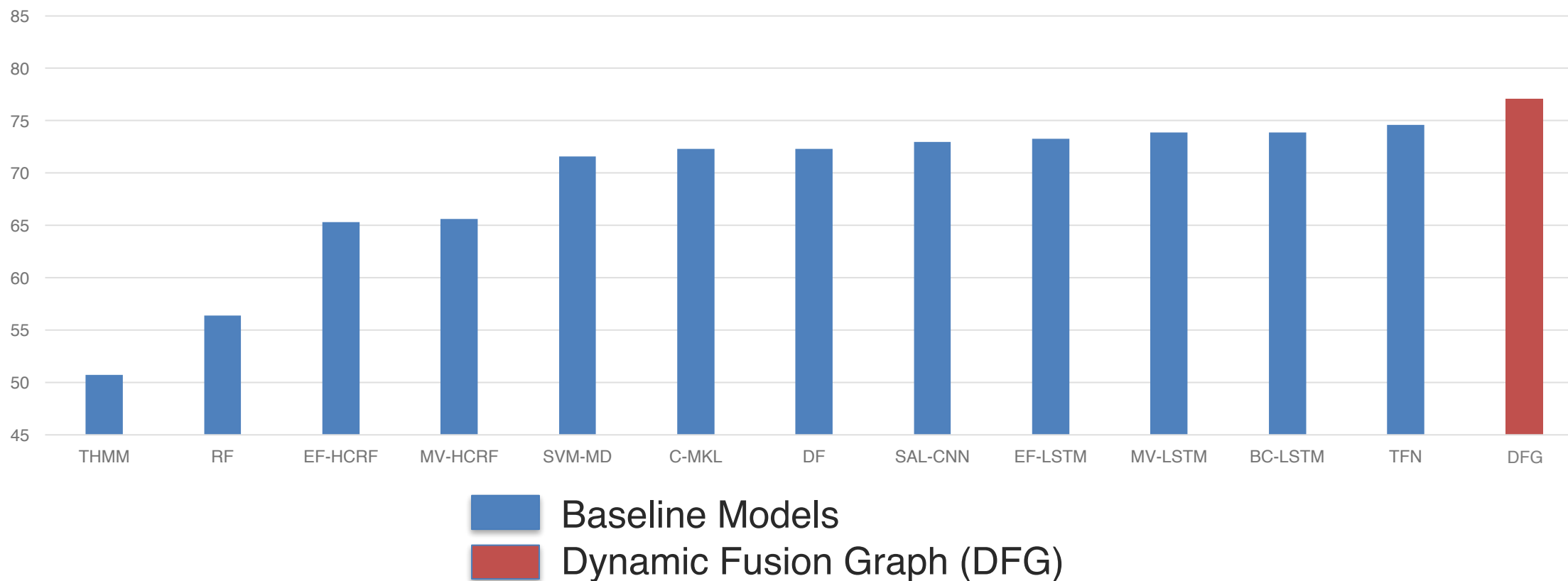

multimodal representation
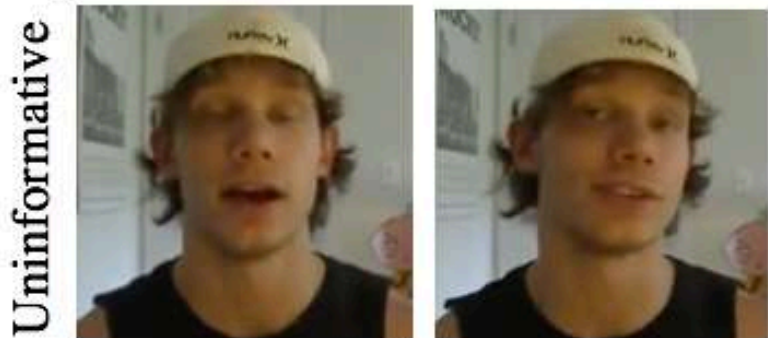
trimodal

bimodal

unimodal

State-of-the-art Results

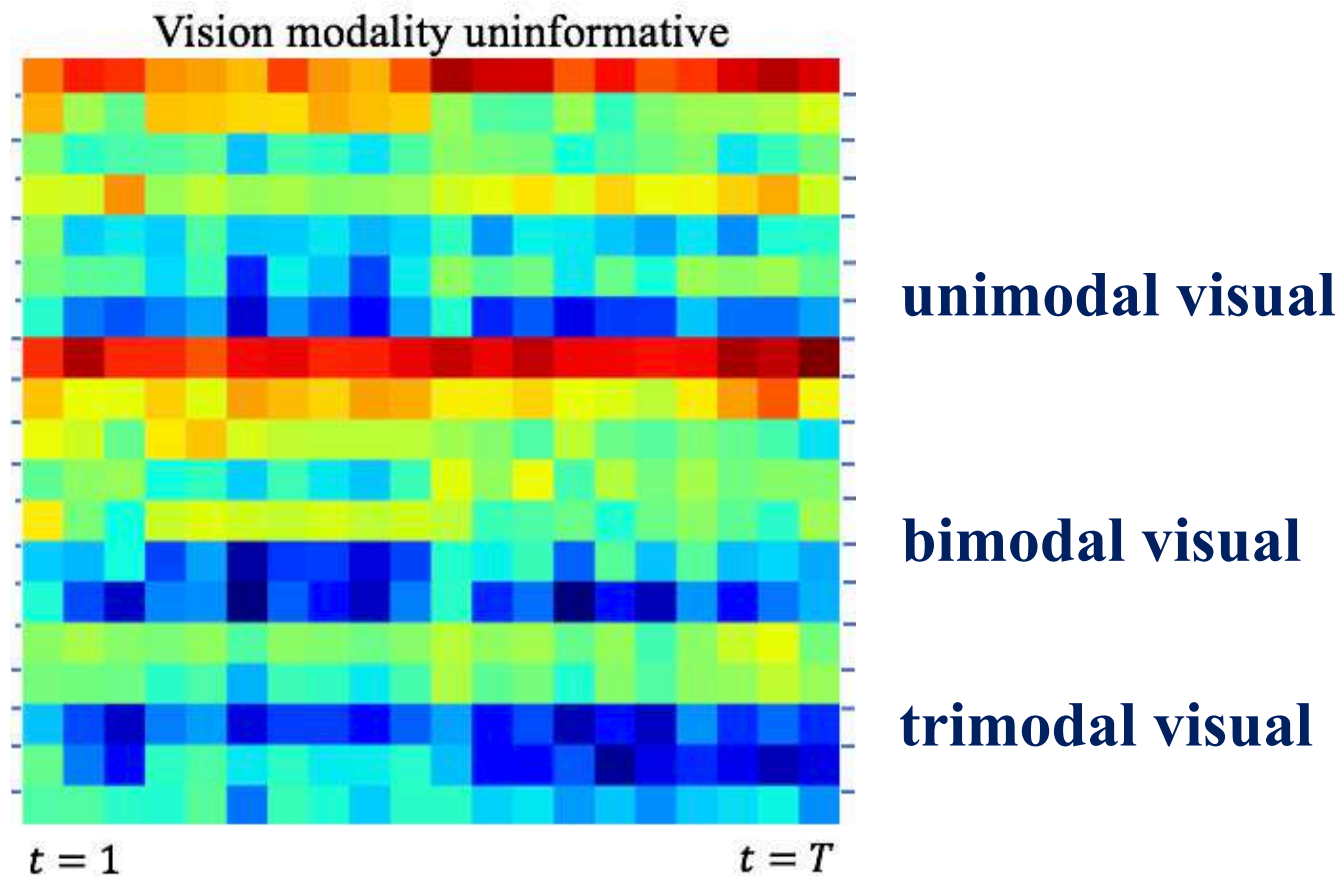## MOSEI Multimodal Sentiment Analysis

Interpretable Fusion

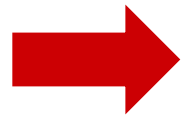Too much too fast, I mean we basically just get introduced to this character...

Uninformative

(angry voice)

Vision modality uninformative

unimodal visual

bimodal visual

trimodal visual

$t = 1$ $t = T$

Computational Modeling of Multimodal Language

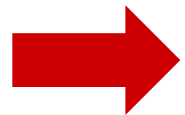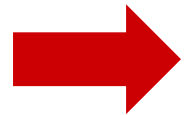**(1)** **MOSEI Dataset**

➤ Diversity in samples, topics, speakers and annotations

(2) **Dynamic Fusion Graph**

➤ State-of-the-art results

➤ Interpretable fusion

The End!

Email: pliang@cs.cmu.edu

Workshop @ ACL 2018
First Workshop and Grand Challenge
on Computational Modeling of Human Multimodal Language
multicomp.cs.cmu.edu/acl2018multimodalchallenge/