# Modeling Mutual Influence of Interlocutor Emotion States in Dyadic Spoken Interaction

Chi-Chun Lee, Carlos Busso, Sungbok Lee, Shrikanth Narayanan
Speech Analysis and Interpretation Laboratory (SAIL)
Viterbi School of Engineering, University of Southern California, Los Angeles, CA 90089
{chiclee,busso,sungbok}@usc.edu, shri@sipi.usc.edu

## Introduction

- Mutual influence - a person's influence on his/her interacting partners behaviors and user states, is evident in multi-persons interaction [1]
- Dyadic interaction can be viewed as two interacting dynamical state systems
- A Dynamic Bayesian Network Model (DBN) is proposed to capture both the temporal evolution and mutual influence of interlocutor emotion states

## Goal

Obtain better emotion recognition performance and bring insights into mutual influence behaviors in dyadic interaction through explicitly modeling **Cross-Speaker Dependency** and **Temporal Dynamics** of emotion states

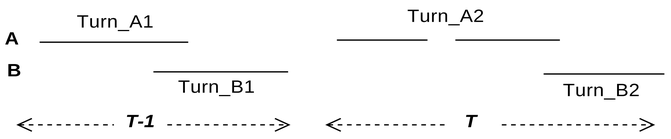
## Database

### Database Description

IEMOCAP Database [2]

- More expressive elicitation of emotion through the setup of dyadic interaction
- Motion captured and audio recorded
- 10 subjects (five pairs of male-female), all actors
- Three categorical emotion labels and two annotation of dimensional attributes on emotion
- 151 dialogs in total

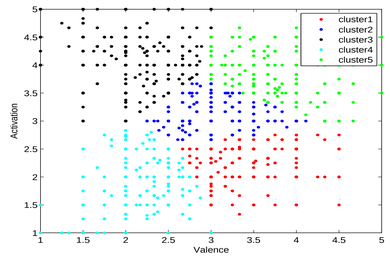
### Analysis Frame Definition




A,B : Interaction Participants
*Turn* is defined as speech portion belonging to each subject
Simultaneous audio information from both speakers

### Emotion Annotation

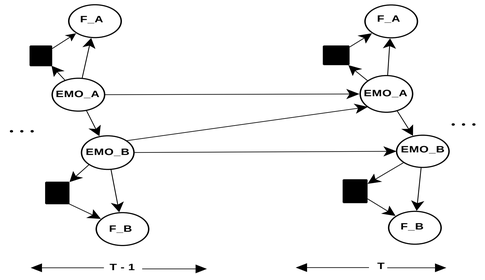
- Dimensional representation of emotion (Valence - Activation)
- Averages taken at *turn* level over utterances and two evaluators
- K-Means Clustering ($k = 5$) to cluster dimensional emotion space ($5^2 = 25$) into five distinct emotion classes
- A total of 8343 *turns*




- "Black" includes 70% of majority-voted *angry* utterances
- "Blue" includes 51% of majority-voted *neutral* utterances

## Proposed Dynamic Bayesian Network

- First-Order Markov Process (Temporal Dynamics)
- Cross Speaker Dependency (Mutual Influence)
- Gaussian Mixture Model on observation feature




## Experiment Setup

**Experiment I :** Recognize 5-Emotion Class
**Experiment II :** Recognize Valence & Activation Separately

- Activation & Valence is each clustered into three classes : High, Medium, Low

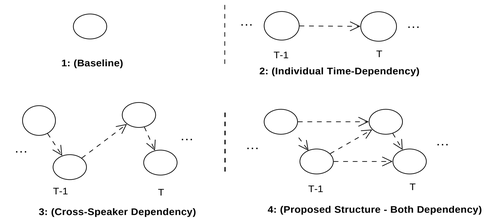
### Feature Extraction

- F0 Frequency
- Energy
- Harmonic to Noise Ratio (HNR)
- MFCC (13 coefficients)
- MFB (27 filter bank coefficients)
- Speech Rate (phonme per second)

Statistics calculated at the *turn* level - 116 dimension feature
*z-normalization* of features with respect to individual's neutral utterances
Forward feature selection using accuracy as stopping criterion

### Emotion Dependency Structure

- Four testing cases on emotion states dependency structure
- Same GMM parameters for all four testing cases
- GMM mixture number = 4
- 15-Fold cross validation ( 140 dialogs / 10 dialogs)




## Experiment Result

| DBN Structure | I: 5 - Emotion Classes | II: Activation-Only (3-Class) | | II: Valence-Only (3-Class) | |
|---|---|---|---|---|---|
| | | *Same* | *Optimized* | *Same* | *Optimized* |
| Chance | 24.29% | 37.11% | 37.11% | 39.21% | 39.21% |
| Baseline - GMM | 51.53% | 62.30% | **63.45%** | 56.59% | 59.89% |
| Time Dependency | 52.68% | 62.02% | 61.92% | 59.78% | 63.40% |
| Mutual Influence | 53.37% | **62.52%** | 62.30% | 59.60% | 62.67% |
| Proposed Model | **55.20%** | 62.35% | 62.49% | **61.26%** | **65.02%** |

- Overall 3.57% absolute (7.12% relative) improvement over baseline performance in Experiment I
- Speech related features performed better with Activation dimension without emotion dependency structure
- Time dependency & Mutual influence both provide improvement of accuracy over baseline
- Valence dimension seems to benefit more from this modeling

## Conclusion and Future Work

- By incorporating context and mutual influence of interlocutor emotion states, we can obtain track emotion states more reliably
- Limitation: Dimension attribute annotation, and single modality cues
- Mutual influence can happen at multiple levels of human-human interactions
- Application including dialog analysis of real life data

## References

[1] J. K. Burgoon, L. A. Stern, and L. Dillman, *Interpersonal Adaptation: Dyadic Interaction Patterns*. Cambridge University Press, 1995.

[2] C. Busso, M. Bulut, C.-C. Lee, A. Kazemzadeh, E. Mower, S. Kim, J. Chang, S. Lee, and S. Narayanan, "IEMOCAP: Interactive emotional dyadic motion capture database," *Journal of Language Resources and Evaluation*, vol. 42, pp. 335–359, 2008.

## Acknowledgements

This research was supported in part by funds from NSF, Army, and USC Annenberg Fellowship