Affective State Recognition in Married Couples' Interactions Using PCA-based Vocal Entrainment Measures with Multiple Instance Learning

Chi-Chun Lee, Athanasios Katsamanis, Matthew P. Black, Brain R. Baucom, Panayiotis G. Georgiou, Shrikanth S. Narayanan
http://sail.usc.edu

Introduction

1. Behavioral Signal Processing

- Signal processing and machine learning to sense/ analyze/model human behavior
- Overarching goal: Jointly exploit expert and data-driven, computational model-based knowledge to inform theory and practice in the domains of behavioral sciences

2. Affective State Recognition by Vocal Entrainment

- Affective ratings are provided by experts following the Social Support Interaction Rating System
- Classification is based on novel vocal entrainment measures and is achieved by multiple instance learning

Motivation

1. Multiple Instance Learning

- A saliency-based prediction model
- A natural fit for analyzing available Couples' Therapy Data (Longitudinal study by UCLA/UW)
- Session-level coding is available but no micro-coding

2. Quantification of Vocal Entrainment as Features

- Signal-derived quantification based on PCA [1]
- Retain interpretability of the feature space

Database

1. Couples' Therapy Database [2]

- 10-minute long video recordings of distressed couples' discussion
- 569 sessions of problem-solving interactions
- Global Positive & Global Negative session-level affect ratings for each spouse per session
- 280 sessions of top 20% of affect ratings (*high-positive* & *high-negative*)
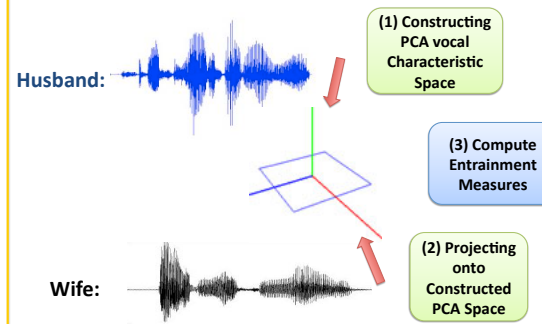
Representative Vocal Features

1. Prosodic Features

- Pitch (4x2) : third order polynomial fit
- Intensity (2x2): mean & variance
- Speech Rate (2x1): word duration

2. Spectral Features

- MFCCs (2x30): mean and variance

Quantifying Vocal Entrainment

Husband:

(1) Constructing PCA vocal Characteristic Space

(3) Compute Entrainment Measures

(2) Projecting onto Constructed PCA Space

Wife:

Experiments Setup

1. Classification Setup

- **Multiple Instance Learning :** EM-Diverse Density
- Classification is based only on (toward, from) vocal entrainment features for each instance (turn)
- Leave-one-couple out cross-validation
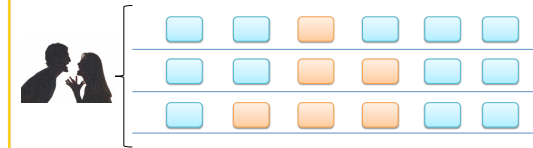
2. Experiment I – Parameter Tuning

- **Test :** mutual information + maximum accuracy

3. Experiment II – Different Instance Lengths

- **Test :** Single, Double, and Triple

MIL with Vocal Entrainment

salient **vocal entrainment over multiple turns to predict the session-level affective rating**

Results of Experiments

1. Results of Experiment I

Classifier	Accuracy (%)
Chance	50.00%
Baseline	51.79%
Majority Vote	44.64%
One$_{maxA}$	50.71%
Proposed	**53.93%**

2. Results of Experiment II

Classifier	Accuracy (%)
Single	47.5%
Double	45.36%
Triple	47.86%
Proposed	**53.93%**

3. Extensions

- Factorial HMM with improved PCA-based Entrainment Quantification (**62.8%**)

References

[1] C.-C. Lee, M. P. Black, A. Katsamanis, A. C. Lammert, B. R. Baucom, A. Christensen, P. G. Georgiou, and S. S. Narayanan, "Quantification of prosodic entrainment in affective spontaneous spoken interactions of married couples," in Proceedings of Interspeech,2010.

[2] A. Christensen, D. Atkins, S. Berns, J. Wheeler, D. H. Baucom, and L. Simpson, "Traditional versus integrative behavioral couple therapy for significantly and chronically distressed married couples," J. of Consulting and Clinical Psychology, vol. 72, pp.176–191, 2004.

[3] Q. Zhang, S. Goldman, "EM-DD: An Improved Multiple Instance Learning Technique", NIPS 2001