A Signal-based Approach to Compute Vocal Behavioral Matching with Application to Interaction Dynamics Analysis

Chi-Chun Lee (Jeremy), Panayiotis G. Georgiou, Shrikanth S. Narayanan
Signal Analysis and Interpretation Laboratory,
University of Southern California

Work Supported by National Science Foundation, the Department of Defense, and the National Institute of Health.

Introduction

1. Behavioral Matching/Entrainment

- Naturally-spontaneous coordination between interacting dyads' behaviors at multiple levels across multiple communicative channels
  - Interaction synchrony, accommodation, mirroring – in general a felt sense of "**in-sync**"
- Goal: compute the *degree* of **vocal** entrainment ('speaking style' matching)
- Implication: affective mechanism, communication efficiency, children's learning of social and communicative skills

2. Computational Challenges

- Subtle phenomenon: difficult for human annotation
- Turn-taking structure of human conversations
- Multiple informative vocal cues
- Inherent directionality of entrainment process
- Lack of extensions from classical synchrony measures

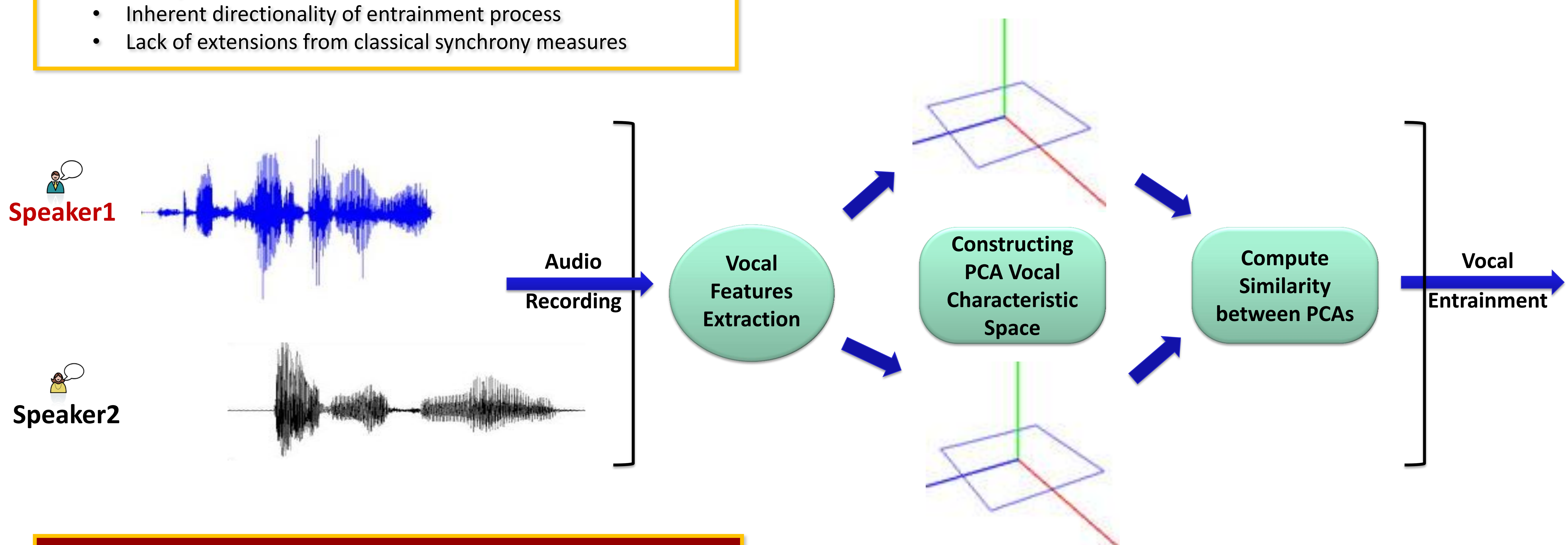
Computational Scheme

1. Quantification Scheme

- Compute similarity measures between PCA-represented vocal characteristics of the interacting dyad at each speaking turn
- Work flow is demonstrated in Figure below

2. Steps to Quantify Vocal Entrainment

- Extract vocal features with robust speech processing technique
- Parameterize vocal features at each word level
- Perform PCA to represent vocal characteristic spaces
- Compute similarity measures as quantitative descriptors of vocal entrainment between the two PCA's



Representative Vocal Features

1. Explicit Speaking Style

- Fundamental frequency ~ intonation
- Intensity / Energy ~ loudness
- Syllable rate ~ speech rate

2. Implicit Speaking Style

- Mel-Frequency Cepstral Coefficients (MFCC)

3. Acoustic Feature Parameterization

- Legendre Polynomial Fit w/ Duration
- Statistical Functions

Case Study: Couple Therapy

1. Couples Therapy Database [1]

- Audio-video recording of real and distressed couples' interactions (134 unique couples)
- 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*) for classification

Similarity Computation [2]

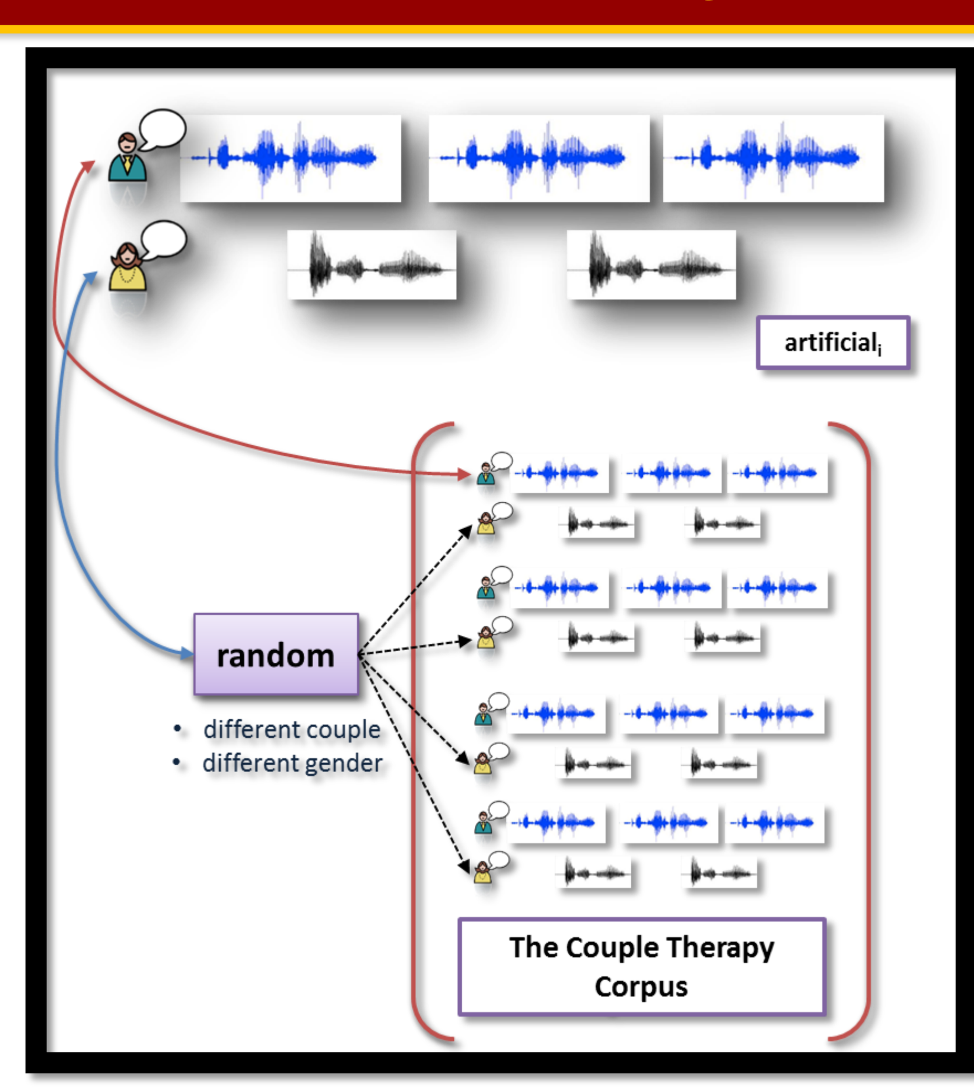
1. Symmetric Metrics

- Sum of cosine angles between two sets of PCA components from two time series

2. Directional Metrics

- Kullback-Leibler divergence (KLD) between two normalized variance vectors as: computed as representing

Statistical Analysis



1. Compare 'Artificial' Dialogs to 'Real' Dialogs

- Vocal entrainment is higher in real dialogs
- Natural cohesiveness in spontaneous dialogs

2. Affect Recognition Accuracy

- Baseline Chance = 50.00%
- Factorial Hidden Markov Model: 62.89% [2]

Reference

[1] 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.

[2] Chi-Chun Lee, Athanasios Katsamanis, Brian Baucom, Matthew P. Black, Andrew Christensen, Panayiotis Georgiou and Shrikanth Narayanan, Computing Vocal Entrainment: A Signal-derived PCA-based Quantification with Application for Affect Analysis in Married Couples' Interactions, in: Journal of Computer Speech and Language, 2012 (in press)