An Analysis of PCA-based Vocal Entrainment Measures in Married Couples' Affective Spoken Interactions

Chi-Chun Lee[1], *Athanasios Katsamanis*[1], *Matthew P. Black*[1],
Brian R. Baucom[2], *Panayiotis G. Georgiou*[1], *Shrikanth S. Narayanan*[1,2]

[1]Signal Analysis and Interpretation Laboratory (SAIL), Los Angeles, CA, USA
[2]Department of Psychology, University of Southern California, Los Angeles, CA, USA
http://sail.usc.edu

Abstract

Entrainment has played a crucial role in analyzing marital couples interactions. In this work, we introduce a novel technique for quantifying vocal entrainment based on Principal Component Analysis (PCA). The entrainment measure, as we define in this work, is the amount of preserved variability of one interlocutor's speaking characteristic when projected onto representing space of the other's speaking characteristics. Our analysis on real couples interactions shows that when a spouse is rated as having positive emotion, he/she has a higher value of vocal entrainment compared when rated as having negative emotion. We further performed various statistical analyses on the strength and the directionality of vocal entrainment under different affective interaction conditions to bring quantitative insights into the entrainment phenomenon. These analyses along with a baseline prediction model demonstrate the validity and utility of the proposed PCA-based vocal entrainment measure.

Index Terms: vocal entrainment, couples therapy, behavioral signal processing, principal component analysis

1. Introduction

In a dyadic spontaneous spoken interaction, the interlocutors exert mutual influence on each other's behaviors. This mutual influence on the dyad's behaviors guides the dynamic flow of the interaction. It is in this context, the term - *interaction synchrony*, a.k.a *entrainment*, is used to describe the phenomenon of a naturally occurring coordination between interacting individuals' behaviors both in timing and form. There has been research works on attempting to quantify specific entrainment behaviors, such as voice activity rhythm [1] and gestures [2], and they have shown that it is essential to apply quantitative methods for analyzing interpersonal interaction dynamics in fine details. Entrainment in conversation describes an important aspect of human interaction dynamics, since it is believed that variations in the pattern of entrainment phenomenon can offer insights into the behaviors of the interacting individuals; this is especially critical in understanding interaction patterns when the underlying behavior is deemed atypical or distressed. This has inspired the investigation of new computational approaches, referred as behavioral signal processing (BSP), to problems in mental health such as couples therapy, addiction behavior, depression, and autism spectrum disorder diagnosis/analysis. The aim of BSP is to automatically analyze abstract human behaviors/states from low level signal measurements such as from audio and video recordings of interactions. In this work, we attempt to quantify *vocal* entrainment in the spoken interactions of married couples engaged in affective problem-solving sessions during marital therapy using such signal processing techniques. A major motivation for this quantitative study of vocal entrainment comes from various psychological studies that have stated the importance of entrainment phenomenon in understanding the nature of couples' interactions [3].

Across a variety of research domains, e.g., econometrics, neuroscience, physical coupled system studies, etc, a long list of *synchronization* measures [4] have been utilized to quantify interdependence between time series and associated variables. These measures often lack straightforward methods to handle complex interaction scenarios like human-human conversations, where the analysis window length (e.g., length of each speaking turn) per channel (e.g., a speaker in the conversation) varies across time and speakers; the signals associated with human conversations can also be very abstract and complex. The two variables in the time series (corresponding to the interlocutors in the dyad) do not occur simultaneously because of the inherent turn-taking structure of human conversations. These phenomenon often violate the underlying assumption when applying classical synchronization measures on signals of interest. Furthermore, majority of these measures are symmetric measures that do not provide information on the directions of synchronization.

In order to improve upon our previous work [5] of quantifying vocal entrainment, we incorporate an expanded list of vocal features and derive a new quantitative vocal entrainment measure based on Principal Component Analysis (PCA). In this work, we propose the quantification of vocal entrainment as the amount of variability preserved when representing a speaker's (say, SP1) vocal characteristics in the vocal characteristics space of another speaker (say, SP2). The vocal characteristics space is constructed using PCA with acoustic cues. Intuitively, the larger the amount of variability preserved, the higher the vocal entrainment level. This method can address both of the aforementioned concerns because of its utilization of projecting vocal features onto the transformed vocal characteristic subspace for any variable length of speech features. Furthermore, for a given speaker pair (say SP1, SP2) in an interaction, this method can generate two directions of vocal entrainment when we look at any single speaker, say SP1: one corresponds to how much SP1 is entraining *toward* SP1, and the other corresponds to how much SP1 is getting entrained *from* SP2.

Various psychology research studies [6, 7] and our own previous work [5] indicate that the general existence of a higher level of entrainment when a spouse is rated as having positive affect compared to rated as having negative affect. While the relationship between entrainment phenomenon and emotion can be complex [8], we rely on this general trend to investigate the use of the proposed PCA-based vocal entrainment measures. Our analysis on the directionality of entrainment further indicates that when a spouse is rated as having positive affect, he/she shows statistically significant more vocal entrainment *toward* his/her interacting partners, but is not eliciting entrainment *from* his/her interacting partners. Finally, we use support vector machine (SVM) to design a baseline prediction model for classifying session-level code of *high positive* vs. *high negative* affect on each spouse using *only* vocal entrainment measure.

The paper is organized as follows: we describe the database and research methodology in section2. Experiment setup and results are in section 3, and conclusions are in section 4.

2. Research Methodology

2.1. Database

The data that we are using was collected as part of the largest longitudinal, randomized control trial of psychotherapy for severely and stably distressed couples [9]. The database consists of audio-visual data recordings: a single channel far-field microphone, split screen videos, and observation coding on the behaviors of these real married couples. Multiple trained evaluators were instructed to code the behaviors of each spouse using the two standard manual codings, the Social Support Interaction Rating System (SSIRS) and the Couples Interaction Rating System (CIRS), resulting in 33 session-level codes for each spouse on their interaction. There are a total of 569 sessions (117 unique couples) of couples engaging in problem solving interactions in which an issue in their relationship was raised and discussed. Since the manual transcripts are available, the audio data was automatically segmented into pseudo-turns (with speaker identification: husband, wife, unknown) and aligned to the word transcripts using a software, SailAlign [10]. These pseudo-turns are considered as *speaking turns* in this research work because they correspond to the speaking portion of the same speaker before the other speaker takes over the floor. The audio data qualities vary a lot from session to session; therefore, we use only a subset of 372 sessions out of 569 sessions because they meet the criteria of 5 dB SNR and 55% speaker segmentations after this automatic process. Details of the database can be found in the previous work [11].

The focus of this work is to quantitatively examine the vocal entrainment of married couples in sessions where either spouse was rated with *extreme* affective states (positive & negative). The emotional rating is the code “Global Positive Affect” and “Global Negative Affect” (based on SSIRS) on each spouse at the session level. We focus on those sessions out of the 372 sessions that spouse was rated in top 20% of positive and negative emotion on the sessions and denote them as being *high positive* emotion and *high negative* emotion in this work. Based on this selection of extreme affective states, it results in a total number of 280 sessions with 81 unique couples of which 140 sessions correspond to *high positive* emotion and another 140 sessions correspond to *high negative* emotion to be used in this work.

2.2. PCA-based Vocal Entrainment Measures

The core idea behind this quantification of vocal entrainment is to construct a basis set representing *speaking characteristics* space of an interlocutor per speaking turn using PCA. The entrainment level is essentially defined as a measure of similarity when projecting another interlocutor’s speaking characteristics onto this constructed space of speaking characteristics; in this case, the metric is the amount of preserved variance of vocal features from one interlocutor while projecting onto the other interlocutor. A schematic example of how to compute the two directions (*toward*: ve_{TO}, *from*: ve_{FR}) of vocal entrainment for an interlocutor’s, husband, speech turn, H$_i$, in an married couple interaction is shown in Figure 1. The steps listed below are used to compute the husband’s ve_{TO} at turn H$_i$:

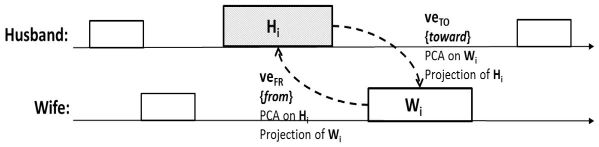

Figure 1: Example of Computing Two Directions of Vocal Entrainment for Turns H$_i$

1. Extract appropriate vocal features, X_1, to represent husband’s speaking characteristics at turn H$_i$.
2. Perform PCA on z-normalized of X_1, such that $Y_1^T = D_1 X_1^T$.
3. Predefined a variance level ($v_1 = 0.95$) to select L-subset of basis vectors, D_{1L}.
4. Project the z-normalized vocal features, X_2 extracted from wife’s speech at turn W$_i$, using D_{1L}.
5. Compute vocal entrainment measure as the ratio of represented variance of X_2, in W_{1L} basis, and the predefined variance level in step 3.

We can compute the other direction of entrainment, ve_{FR}, by interchanging X_1 with X_2. There are two major motivations behind these PCA-based vocal entrainment measures. First is the elimination of concerns associated with imposing heuristics in the computation of conventional synchronization measures due to the the turn-taking structures of human conversation and variable length of speaking turns (resulting in different number of vocal feature vectors sequences per speaking turn). These two factors can raise concerns on the reliability of using classical measures. However, with the PCA-based measures, because the representation resides in another transformed space, the issues of non-simultaneously occurring time series and variable-length analysis chunks are both lessened. Second is the ability to introduce the directionality of entrainment at each speaking turn per speaker. As we can see from Figure 1, there can be two directions of vocal entrainment for a given spouse at each of his/her speaking turn. This directionality can be important to understand the details of entrainment phenomenon.

2.3. Representative Vocal Feature Set

The method describes in Section 2 relies on an appropriate set of acoustic features to represent speaking characteristics. In order to capture the dynamics of the speaking characteristics, the PCA is done on a speaking turn where the vocal features are computed at the word level. There are two different categories of vocal features used in this work: prosodic features and spectral features. The details of the raw acoustic extractions from audio files with necessary preprocessing and speaker-normalization are described in previous work [11]. The following is the list of final set of acoustic features calculated per word (resulting from automatic alignment) to represent the speaking characteristics.

- Prosodic Features (Pitch x4) : third-order polynomial fit on the pitch contour per word.
- Prosodic Features (Energy x2) : mean and variance of the energy per word.
- Prosodic Features (Word Duration x1): the word duration.
- Spectral Features (MFCC 2x15): mean and variance of 15 dimensional MFCC per word.

This list combined with the first order delta features generates a 74-dimensional (37x2) vocal feature vector per word. Depending on the length of speaking turns, it would result in a variable length of 74-dimensional vocal feature sequences. PCA are performed on the *merged-turns*, (merging speaking turns into *merged-turn* such that it has least 74 samples), in order to generates a unique set of basis vectors.

3. Experiment Setups & Results

Three different experiments were set up to analyze different aspects of this PCA-based vocal entrainment measure.

- **Experiment I:** To investigate if the proposed PCA-based vocal entrainment measures offer a reasonable quantification of vocal entrainment using two different hypothesis testings.
- **Experiment II:** To analyze the direction of the PCA-based vocal entrainment under different conditions of affective married couples' interactions.
- **Experiment III:** To discriminate affective state using this PCA-based vocal entrainment measure as features with Support Vector Machine.

3.1. Experiment I

We used two different approaches in verifying that the PCA-based vocal measure is indeed a viable quantitative measure of vocal entrainment. First, we rely on the fact that there has been a general understanding that when couples are engaged in an interpersonal interaction with a more *positive* emotion, the entrainment level is expected to be higher than with a *negative* emotion. The second is to show that the entrainment measures computed this way between interacting couples is statistically higher than computing the entrainment measure between random pair of couples not engaged in conversations.

3.1.1. Hypothesis Testings Setup & Results

The first hypothesis testing was to verify the approach of PCA-based vocal entrainment by using the Student's T-Test ($\alpha = 0.05$) to examine whether the value is bigger in cases when a spouse was rated with *high positive* emotions compared with *high negative* emotions. The distribution of the PCA-based entrainment measures was approximately normal. Table 1 shows the results of the hypothesis testing.

Table 1: *Entrainment Levels: Higher in Positive Emotion vs Negative Emotion.*

Entrainment Type	*High* Positive	*High* Negative	p-value
Toward (ve_{TO})	0.8276	0.8198	**0.0103**
From (ve_{FR})	0.8307	0.8256	0.0699

Table 1 shows that when a spouse was rated with positive affective state, the associated PCA-based entrainment measures are higher for both of the direction (*toward* and *from*) though only the direction of *toward* passed the ($\alpha = 0.05$) significance level. This result provides an evidence that the PCA-based entrainment measure describes the entrainment phenomenon that is generally understood in marital communication.

Another hypothesis testing was conducted using the Student's T-Test ($\alpha = 0.05$) to examine whether this PCA-based entrainment computed in sequence of turn takings for actual interacting couples has a larger value than when computed for any random pair of speaking turns. The intuition is that if this method captures the notion of coherence in dialogs, this measure should have a higher value compared to when computing two turns that are randomly selected (between two people that were not engaged in direct interaction). Instead of examining both directions separately, average of the values were computed across all 372 sessions (not restricting to only positive vs. negative sessions). Random entrainment values were computed with 10,000 random draws with replacement of a pair of turns from non-interacting couples. Table 2 shows the statistical testing result.

Table 2: *Entrainment Levels: Higher in Pairs of Sequence in Turn-Taking vs. Random Pair of Turns.*

	Pairs of Turns	Random Pairs	p-value
Avg. of Entrainment	0.8266	0.8231	**0.018**

Table 2 provides additional corroborating statistical evidence that indeed this PCA-based method of computing entrainment captures a notion of vocal synchronization because the value is greater overall when we compute it across turn-sequences of interacting couples compared with turn-pairs of non-interacting "couples". These two hypothesis testing experiments provide some grounding evidence that the signal-derived PCA-based vocal entrainment measure is a viable method to quantify interpersonal synchronization.

3.2. Experiment II

In Experiment II, we extend our statistical analysis to analyze the strength of vocal entrainment in each interaction direction (*toward* and *from*) given different conditions, termed here *interaction atmosphere*. Here, for our problem context, we define interaction atmosphere as three types: 1, both spouses were rated as having *high positive* emotion, 2, only one spouse was rated with *high positive* emotion or with *high negative* emotion, and 3, both spouses were rated as having *high negative* emotions. The following is the list of the statistical testings with the Student's T-Test ($\alpha = 0.05$).

- **Test 1:** Comparison of entrainment measure for type 1 vs. type 3 interactions: alternative hypothesis states that the entrainment values are higher in type 1.
- **Test 2:** Comparison of entrainment measure for type 2 interactions: alternative hypothesis states that entrainment values are higher when one spouse was rated as *high positive* vs. one spouse was rated as *high negative*.
- **Test 3:** Comparison of entrainment measures for type 1 vs. type 2 interactions: alternative hypothesis states that when both spouses were rated as *high positive*, entrainment values are higher compared to when only one spouse was rated as *high positive*.
- **Test 4:** Comparison of entrainment measure for type 1 vs. type 2 interactions: alternative hypothesis stating that when both spouses were rate *high negative*, entrainment values are lower compared with only one spouse was rated as *high negative*.

The summary of the statistical testing results of Experiment II is in Table 3. Several notable points can be made with the result in Table 3. First, the vocal entrainment measures are higher (in both directions) where both spouses were rated as having *high positive* emotions (Test 1), which is expected as suggested by various psychology literatures. Second, with this quantification of vocal entrainment, results suggest that when one spouse was rated with *high positive* emotion, he/she shows higher values of entrainment *toward* his/her interacting partner compared to when he/she was rated as *high negative* (Test 2). This implies that when a person is in a more positive emotion, his/her

Table 3: *Hypothesis Testing Summary (Various Interaction Atmosphere Types.*

Test # (Entrainment Type)	Mean of H_o	Mean of H_a	p-value	Test # (Entrainment Type)	Mean of H_o	Mean of H_a	p-value
Test 1 (*toward*)	0.8196	0.8289	**0.0314**	Test 1 (*from*)	0.8196	0.8289	**0.0314**
Test 2 (*toward*)	0.8189	0.8265	**0.050**	Test 2 (*from*)	0.8311	0.8321	0.3831
Test 3 (*toward*)	0.8265	0.8289	0.3126	Test 3 (*from*)	0.8289	0.8321	0.7741
Test 4 (*toward*)	0.8189	0.8196	0.5635	Test 4 (*from*)	0.8311	0.8196	**0.009**

vocal characteristics are becoming similar toward his/her interacting partner to possibly ease the tension of the interaction or provide support. However, the results indicate that his/her interacting partners may not have displayed such entrainment toward him/her. Test 3 results suggest that there is no difference in the level of vocal entrainment when both spouses were rated with positive emotion compared to when only one spouse was rated with positive emotion. Lastly, this results suggest that when both spouses were rated as *high negative*, they receive less vocal entrainment from their interacting partner compared to when only one spouse was rated as *high negative* (Test 4). This outcome is also intuitive because when couples are both negative, they would be less willing to entrain toward one another (less likely to provide emotional support to each other). Through this series of statistical testings, it is encouraging to observe that this method can be a viable approach to perform detailed analysis of entrainment in relation to psychologist's affective rating of these distressed couples with a potential of performing many more testings on entrainment for various interaction conditions.

3.3. Experiment III

The goal of this experiment is to study the predictive ability of the vocal entrainment measure in recognizing spouse's session-level affective codes. We performed a baseline binary classification using Support Vector Machine (with radial basis functions) to differentiate *high positive* vs. *high negative* affective states using this vocal entrainment measure. We focus on only one direction of the entrainment (*toward*) for each spouse, since in Table 3, it exhibits statistical significance difference between *high positive* and *high negative* affective states. Nine different statistical functionals were computed per session (mean, variance, range, maximum, minimum, 25% quantile, 75% quantile, interquartile range, median). Evaluation was done using an leave-one-couple-out cross validation, and we obtained an recognition rate of 51.79%. A more detailed classification setup of recognizing affective state using a multiple instance learning framework further improves recognition rate to 53.93% with salient vocal entrainment measures [12].

4. Conclusions & Future Works

The entrainment phenomenon is an integral aspect when analyzing couples interactions. Computational measures of vocal entrainment can provide a quantitative characterization accompanying qualitative descriptions of this natural human communication phenomenon. In this work, we propose a PCA-based vocal entrainment measure. It relies on the idea that to effectively capture this subtle similarity between an interacting dyad, we first construct a space (PCA) representing speaking characteristics of each interlocutor with a set of common acoustic features; then, the entrainment level is computed as the preserved variability of another speaker represented in the transformed-feature space of the original speaker. Analysis presented in Section 3 shows that this is indeed a viable approach to quantify vocal entrainment, and various statistical analyses using real couple interaction data have shown the differences in the strength of directionality of vocal entrainment when each spouse is rated with *high positive* compared to *high negative* affect.

Future works includes investigation of better representation of speaking characteristics using various acoustic cues since a suitable representation of the speaking style is a crucial aspect while utilizing this method of PCA-based entrainment. Another research direction involves utilizing a more sophisticated subspace construction method to overcome inherent problems of PCA, such as its sensitivity to outliers A further direction is to construct different representations that effectively capture nonverbal behaviors. Since entrainment can provide insights into conducting research on human-human communication, we would like to extend this quantification scheme in hope to offer psychology experts another choice of useful objective tools for analysis of married couples communication.

5. Acknowledgments

This research was supported in part by funds from the National Science Foundation, the Department of Defense, and the National Institute of Health.

6. References

[1] A. R. McGarva and R. M. Warner, “Attraction and social coordination: Mutual entrainment of vocal activity rhymes,” *Journal of Psycholinguistic Research*, vol. 32, no. 3, pp. 335–354, 2003.

[2] M. J. Richardson, K. L. Marsh, and R. Schmit, “Effects of visual and verbal interaction on unintentional interpersonal coordination,” *Journal of Experimental Psychology: Human Perception and Performance*, vol. 31, no. 1, pp. 62–79, 2005.

[3] K. Eldridge and B. Baucom, *Positive pathways for couples and families: Meeting the challenges of relationships.* WileyBlackwell, ch. (in press) Couples and consequences of the demand-withdraw interaction pattern.

[4] J. Dauwels, F. Vialatte, and A. Cichocki, “Diagnosis of alzheimers disease from eeg signals: Where are we standing?” *Current Alzheimer's Research (Invited Paper)*, 2011.

[5] 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.

[6] M. Kimura and I. Daibo, “Interactional synchrony in conversations about emotional episodes: A measurement by 'the between-participants pseudosynchrony experimental paradigm',” *Journal of Nonverbal Behavior*, vol. 30, pp. 115–126, 2006.

[7] L. L. Verhofstadt, A. Buysse, W. Ickes, M. Davis, and I. Devoldre, “Support provision in marriage: The role of emotional similarity and empathic accuracy,” *Emotion*, vol. 8, no. 6, pp. 792–802, 2008.

[8] J. M. Gottman, “The roles of conflict engagement, escalation, and avoidance in marital interaction: A longitudinal view of five types of couples,” *Journal of Consulting and Clinical Psychology*, vol. 61, no. 1, pp. 6–15, 1993.

[9] 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.

[10] A. Katsamanis, M. P. Black, P. G. Georgiou, L. Goldstein, and S. S. Narayanan, “SailAlign: Robust long speech-text alignment,” in *Very-Large-Scale Phonetics Workshop*, Jan. 2011.

[11] M. P. Black, A. Katsamanis, C.-C. Lee, A. C. Lammert, B. R. Baucom, A. Christensen, P. G. Georgiou, and S. S. Narayanan, “Automatic classification of married couples' behavior using audio features,” in *Proceedings of Interspeech*, 2010.

[12] C.-C. Lee, A. Katsamanis, M. P. Black, B. R. Baucom, P. G. Georgiou, and S. S. Narayanan, “Affective state recognition in married couples' interactions using pca-based vocal entrainment measures with multiple instance learning,” in *Submitted to ACII*, 2011.