Analyzing Temporal Dynamics of Dyadic Synchrony in Affective Interactions

Zhaojun Yang and Shrikanth Narayanan

Signal Analysis and Interpretation Lab (SAIL), University of Southern California, Los Angeles, CA

zhaojuny@usc.edu, shri@sipi.usc.edu

Abstract

Human communication is a dynamical and interactive process that naturally induces an active flow of interpersonal coordination, and synchrony, along various behavioral dimensions. Assessing and characterizing the temporal dynamics of synchrony during an interaction is essential for fully understanding the human communication mechanisms. In this work, we focus on uncovering the temporal variability patterns of synchrony in visual gesture and vocal behavior in affectively rich interactions. We propose a statistical scheme to robustly quantify the turn-wise interpersonal synchrony. The analysis of the synchrony dynamics measure relies heavily on functional data analysis techniques. Our analysis results reveal that: 1) the dynamical patterns of interpersonal synchrony differ depending on the global emotions of an interaction dyad; 2) there generally exists a tight dynamical emotion-synchrony coupling over the interaction. These observations corroborate that interpersonal behavioral synchrony is a critical manifestation of the underlying affective processes, shedding light toward improved affective interaction modeling and automatic emotion recognition.

Index Terms: interactional/dyadic synchrony, synchrony dynamics, acoustic/gesture synchrony, affective interactions

1. Introduction

Human communication is a dynamical and interactive process that is established on a common ground of achieving the interaction goals and sharing mutual interests of the interaction participants. Such an interactive process naturally requires interpersonal coordination. Notably this often invokes interactional synchrony or behavior adaptation, along various behavioral dimensions of spoken words, speech prosody, body gestures and emotional states [1]. The mutual dyadic behavioral influence controls the dynamical flow of a conversation and characterizes the overall interaction patterns. Automatically tracking the interpersonal synchrony over an interaction and thoroughly understanding its temporal dynamics can bring insights into the study of rich human communication mechanisms and the design of human-machine interfaces.

The phenomenon of interpersonal synchrony in human communication has been well-established qualitatively in the behavioral science and psychology research. An extensive literature in these domains has focused on manually observing and assessing such mutual influence. For example, in the research on interpersonal relations, behavior synchrony in a couple's interaction has been shown to offer predictive markers of the couple's mental distress and well-being conditions [2] [3]. Motivated by such findings, engineering researchers have developed computational approaches to automatically quantify behavioral coordination, and synchrony, for comprehensively characterizing human interaction dynamics, and the underlying mental states. Lee *et al.* proposed a PCA-based scheme to quantify turn-wise vocal synchronization and investigated the relationship between the quantified measure and the underlying affective processes in married couples' interactions. A higher degree of vocal synchrony was found for couples with positive emotions [4]. The analysis in our previous work [5] has also empirically demonstrated that the interpersonal coordination patterns of body language depends on the stances assumed.

These existing works however concentrate on the gross degree of behavior coordination over an entire interaction. Little attention thus far has been paid to study the temporal dynamics of interactional synchrony along various behavioral modalities. Human communication results from a complex interplay of the dynamical processes of expressive behaviors, which naturally leads to the dynamical nature of interactional synchrony. Besides assessing the overall coordination strength, characterizing the corresponding temporal variability is essential for comprehensively understanding and computationally modeling the dynamical flow of human communication [6].

This work aims at uncovering the temporal dynamics of dyadic synchrony *w.r.t.* gesture and vocal behavior in affective interactions. Our goal is two-fold: 1) to investigate how the dynamical patterns of interactional synchrony depend on the global affective states of a conversational dyad; 2) to examine how the interpersonal synchrony and a dyad's emotions are dynamically correlated turn by turn over a conversation. To this end, we propose a statistical scheme to automatically quantify turn-wise behavior synchrony. Compared to the previously developed synchrony measures [4] [7], our metric has the advantages of robustness to individual idiosyncrasies and generalizability to diverse behavior aspects. The analysis of the quantified dynamic synchrony measure relies heavily on the functional data analysis (FDA) techniques that provide a useful mathematical framework for exploring the temporal patterns of time series data. We first apply functional PCA (FPCA) to characterize the temporal variability of dyadic synchrony over an interaction. We find that the temporal synchrony patterns vary depending on the global emotional states of an interaction dyad. We further employ functional canonical correlation analysis (FCCA) to uncover the dynamical association between the behavior synchrony and a dyad's emotions over time. Our analysis results reveal that there generally exists a tight dynamical emotion-synchrony coupling over an interaction. These observations shed light upon the complex nature of affective communication and design of improved emotion recognition.

2. Data Description

We use the USC CreativeIT database for the analysis of synchrony dynamics [8]. It is a freely-available multimodal database of dyadic theatrical improvisations. Interactions are goal-driven, which can elicit natural realization of emotions and expressive multimodal behavior. There are 50 interactions in to-

This work was funded by NSF.

tal performed by 16 actors (9 female). Each interaction has an average length of 3.5 minutes. The audio data of each actor was collected through close-talking microphones at 48 kHZ. A Vicon motion capture system with 12 cameras captured detailed full body Motion Capture data at 60 fps, *i.e.*, the (x, y, z) positions of 45 markers on each actor, as shown in Figure 1(a).

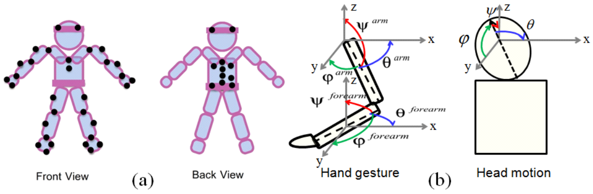

Figure 1: (a) The positions of Motion Capture markers; (b) Euler angles of hand and head joints.

2.1. Gesture and Vocal Acoustic Features

This work considers hand and head gestures which are the most expressive motions in human communication [9]. To extract gesture features, we manually mapped the markers positions (x, y, z) to the joint angles using MotionBuilder [10]. The joint angles are popular for motion animation [11] [12] and gesture dynamics modeling [9] [13]. Figure 1(b) illustrates the Euler angles (θ, ϕ, ψ) of hand (arm and forearm) and head joints in x, y, z directions. The angles of hand and head joints are used as gesture features.

In addition, we represent the vocal behavior by extracting acoustic features of pitch, the rms energy and 12 Mel Frequency Cepstral Coefficients. These features were extracted every 16.67 ms with a 30 ms analysis window to match with the MoCap frame rate. The pitch features were smoothed and interpolated over the unvoiced regions. We further augmented both gesture and acoustic features with their $1st$ derivatives. These extracted features have been shown to be emotion-related and are popular in the affective computing community [5] [14]. All the features were z-score normalized in a subject-dependent way. This work analyzes the gesture and vocal synchrony in a dyad over an interaction using the extracted representations.

2.2. Emotion Annotation

We collected two types of emotion annotations in the database: the time-continuous emotional flow and the global emotional content of an actor in an interaction

2.2.1. Continuous emotion annotation

The time-continuous emotional state of each interlocutor was annotated in terms of the dimensional attributes of activation (excited *vs*. calm) and valence (positive *vs*. negative). Annotators used Feeltrace [15] to continuously indicate the attribute value from -1 to 1 while watching the video [16]. Our work focuses on studying the turn-wise synchrony. Hence, we partition each actor recording into dialog turns according to speech regions. Each turn of a leading interlocutor is paired with the following turn of the partner, resulting in 1296 turn pairs, as illustrated in Figure 2. The activation/valence value in each turn is calculated by averaging the ratings across frames and annotators, and is mapped into low $[-1, 0]$ and high $(0, 1]$ classes. We further define the categorical emotion of a turn pair as low (high) activation/valence if both turns are in the low (high) class. We also define the continuous emotion value of a turn pair by averaging the activation/valence values of both turns.

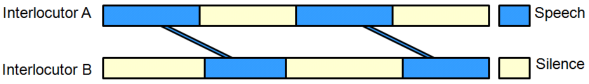

Figure 2: Illustration of setting up dialog turn pairs. Dyadic synchrony is computed for each pair of dialog turns.

2.2.2. Global emotion annotation

The overall perception of activation and valence for each actor in a recording was rated on a 5-point scale by annotators [16]. The global rating values of activation and valence of each actor in an interaction are calculated by averaging the ratings across annotators, and are mapped into low $[1, 3]$ and high $(3, 5]$ classes. The global emotion category of a dyad is defined as low (high) activation/valence if both subjects are in the low (high) class (The cases with both low and high classes will be incorporated in the future analysis).

3. Quantification of Dyadic Synchrony

The goal of this work is to analyze the dynamical changes of behavior synchrony over an interaction. The foremost precondition is to robustly quantify the turn-wise synchrony measure that is generalizable to various behavior aspects. We propose a statistical scheme for this purpose, where the main idea is to represent the behavioral characteristics of an individual at each turn using the subject-independent behavior distribution.

Let $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_T\}$ be a dialog turn of an interlocutor, where T is the frame number and $\mathbf{x}_t \in \mathcal{R}^d$ is a gesture or acoustic feature vector (see Section 2.1) at frame t. We first build a subject-independent model of behavior features based on GMMs to unify information from different subjects. GMMs have shown great success for robust representation modeling in diverse domains [9] [17] [18]. Specifically, the subject-independent model is constructed using the feature vectors in the dialog turns of all the interlocutors: $p(\mathbf{x}|\mathbf{\Theta}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_k, \mathbf{\Sigma}_k)$, where K is the number of mixtures, and $\{\pi_k, \boldsymbol{\mu}_k, \mathbf{\Sigma}_k\}$ are the weight, mean vector and covariance matrix of the k-th component. The parameters $\mathbf{\Theta}$ can be estimated based on the maximum likelihood criterion using Expectation Maximization.

$p(\mathbf{x}|\mathbf{\Theta})$ summarizes the subject-independent generation process of the behavior feature vector $\mathbf{x}$. Our intuition is that incorporating such global generative information in the turn-wise behavior description could bring robustness to the individual idiosyncrasies. Motivated by the work in [5] which has demonstrated that Fisher vector [19] is an effective way of embedding a generative model in behavior description, we describe the behavioral characteristics in the turn $\mathbf{X}$ using Fisher vector: $\mathbf{f}_{\mathbf{x}} = \frac{1}{T}\sum_{t=1}^{T} \nabla_{\mathbf{\Theta}} \log p(\mathbf{x}_t|\mathbf{\Theta})$. $\mathbf{f}_{\mathbf{x}} \in \mathcal{R}^{2dK}$ describes how the parameters $\mathbf{\Theta}$ contribute to the process of generating the local behavior in the dialog turn $\mathbf{X}$. It is a popular and efficient representation of image, audio and video data [20] [21] [22]. In addition to encoding the global statistical information, $\mathbf{f}_x$ provides a unified and generalizable representation form for the turn-taking structure of human conversations where the dialog turns are of variable lengths over time and across interlocutors.

Given a pair of dialog turns $(\mathbf{X}_A, \mathbf{X}_B)$, their behavior characteristics are represented by $(\mathbf{f}_{\mathbf{x}_A}, \mathbf{f}_{\mathbf{x}_B})$. We define the behavior synchrony between the pair of turns based on the angle θ_{AB} between $\mathbf{f}_{\mathbf{x}_A}$ and $\mathbf{f}_{\mathbf{x}_B}$,

$$\sigma = cos^2\theta_{AB},\ cos\theta_{AB} = \frac{\mathbf{f}_{\mathbf{x}_A}^T \mathbf{f}_{\mathbf{x}_B}}{|\mathbf{f}_{\mathbf{x}_A}| \cdot |\mathbf{f}_{\mathbf{x}_B}|}. \quad (1)$$

The cosine similarity is suitable for measuring similarity between high dimensional feature vectors, and has been adopted in the synchrony measure proposed in [4]. σ is bounded in $[0, 1]$. A greater σ value indicates a higher level of dyadic synchrony.

4. Analysis and Results

This section aims at investigating how the quantified synchrony measure dynamically changes over time depending on the emotional states of an interaction dyad.

4.1. Verification of The Synchrony Measure σ

The analysis in this section is to verify the validity of the proposed measure in Equation (1) for characterizing behavior synchrony in human communication. To this end, we compare the computed synchrony measure in each turn pair of the actual interactions with that in any random turn pair. Both turns in a random pair are from different interactions. The assumption under this comparison is that there exists natural mutual behavior influence in human communication. Hence the synchrony degree in a matched turn pair is expected to be higher than that in a random pair. We generate $5,000$ random turn pairs and compute the turn-wise synchrony measure *w.r.t.* both gesture and vocal acoustic features (see Section 2.1).

The t-test comparison reveals that the gesture synchrony in the matched turn pairs (Avg. 0.364) is significantly higher compared to the random pairs (Avg. 0.282) with $p = 0.000$; and that a significantly greater degree of acoustic synchrony is observed in the matched turn pairs (Avg. 0.561) than in the random pairs (Avg. 0.488) with $p = 0.000$. The results support that the proposed synchrony measure is able to appropriately capture the inherent interpersonal behavior cohesiveness in human communication.

In addition, we compare the synchrony measures in distinct turn-pair-wise emotion categories (see Section 2.2.1). As shown in Table 1, both gesture and acoustic synchrony measures of a low-activation turn pair are significantly higher compared to a high-activation pair; a greater degree of behavior synchrony is observed when both interlocutors are rated as high-valence compared to the low-valence ones. This observation is consistent with the well-established fact in the couple therapy studies that couples with positive emotions are more behaviorally coherent in interactions than distressed couples [4] [23], which reinforces the validity of the quantified measure for characterizing the interpersonal mutual influence.

4.2. Synchrony Dynamics in Affective Interactions

Having established on the validity of the quantified synchrony measure, we herein focus on uncovering the temporal dynamics of the turn-level synchrony measure in affective interactions, relying heavily on FDA techniques.

Table 1: Synchrony comparison: low-activation *vs.* high-activation and low-valence *vs.* high-valence.

Synchrony Type	**Low-Activation**	**High-Activation**	p**-value**
Gesture	0.407	0.357	0.0491
Acoustic	0.566	0.530	0.0173
Synchrony Type	**Low-Valence**	**High-Valence**	p**-value**
Gesture	0.302	0.421	0.000
Acoustic	0.517	0.601	0.000

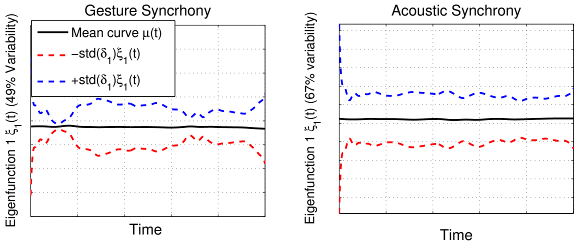

Figure 3: The 1st eigenfunction $\xi_1(t)$ of gesture/acoustic synchrony curves. $std(\delta_1)$ is the standard deviation of the 1st PCA score δ_1 of the synchrony curves along $\xi_1(t)$.

4.2.1. *Functional Data Construction in Interactions*

Functional data analysis (FDA) techniques provide a mathematical framework for exploring the temporal variability of time series data, and are hence useful for uncovering the temporal structure of interactional synchrony. Let $\{y_n\}_{n=1}^N$ be a general time series data. To employ FDA techniques on the data, we first transform the time-discrete data into time-continuous functional data, *i.e.*, $\widetilde{y}(t) = \sum_{k=1}^K c_k\phi_k(t)$, where $\{\phi_k(t)\}_{k=1}^K$ are the predefined basis functions. In this work, we choose B-splines as the basis functions. The coefficients $\mathbf{c}$ are estimated by minimizing the fitting error,

$$\mathbf{c} = \arg\min_{\mathbf{c}} \sum_{n=1}^{N} (y_n - \widetilde{y}(t_n))^2 + \mu \int [\mathcal{D}^2\widetilde{y}(t)]^2. \qquad (2)$$

$\mathcal{D}^2$ denotes the 2nd derivative and defines the roughness of $\widetilde{y}(t)$.

As illustrated in Figure 2, a dyadic interaction is characterized as a sequence of dialog turn pairs $\{(\mathbf{X}_{A,n}, \mathbf{X}_{B,n})\}_{n=1}^N$. In each turn pair $(\mathbf{X}_{A,n}, \mathbf{X}_{B,n})$, we compute a synchrony measure σ_n (see Equation (1)) that is associated with a turn-pair-wise continuous value of activation/valence e_n (see Section 2.2.1), resulting in two one-dimensional time series data, $\{\sigma_n\}_{n=1}^N$ and $\{e_n\}_{n=1}^N$. According to Equation (2), $\{\sigma_n\}_{n=1}^N$ and $\{e_n\}_{n=1}^N$ can be respectively transformed into time-continuous functional data $\widetilde{\sigma}(t)$ and $\widetilde{e}(t)$ that are used in the analysis that follows.

4.2.2. *Temporal Synchrony Dynamics in Relation to Global Dyadic Emotions*

We first investigate how the dynamical patterns of interactional synchrony depend on the global emotions of a dyad. To characterize the temporal variability of dyadic synchrony, we apply FPCA to the functional synchrony data $\widetilde{\sigma}(t)$. FPCA finds the principal variation modes, *i.e.*, a set of orthonormal eigenfunctions $\{\xi_j(t)\}_j$, of the functional data [24]. $\xi_j(t)$ is computed such that the data variance along the direction is maximized: $\xi_j(t) = \arg\max var(\int(\widetilde{\sigma}(t) - \mu(t))\xi_j(t)dt)$, where $\mu(t)$ is the mean curve. The set of eigenfunctions represent the principal components of temporal variation of the synchrony curves. The projection of $\widetilde{\sigma}(t)$ along $\xi_j(t)$ is defined as the PCA score, *i.e.*, $\delta_j = \int(\widetilde{\sigma}(t) - \mu(t))\xi_j(t)dt$, summarizing the overall dynamics of the synchrony along the corresponding eigenfunction.

We apply FPCA respectively to the gesture and acoustic synchrony curves from all the interactions. Figure 3 presents the 1st eigenfunction $\xi_1(t)$ of the gesture/acoustic synchrony curves, which describes the dominant synchrony variability. We can observe distinct variation modes for gesture and acoustic synchrony: the variability of acoustic synchrony is more temporally stable in general, while the gesture synchrony involves more drastic oscillations over time. Figure 4 presents the scatterplots of the 1st and 2nd PCA scores, *i.e.*, δ_1 and δ_2, of ges-

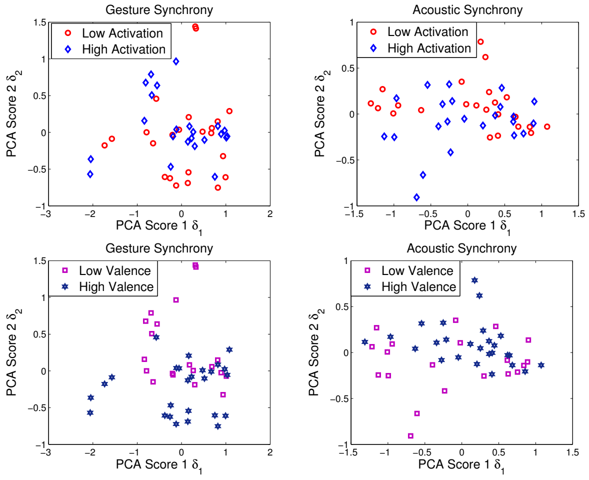

Figure 4: Scatterplots of 1st and 2nd PCA scores of gesture/acoustic synchrony curves.

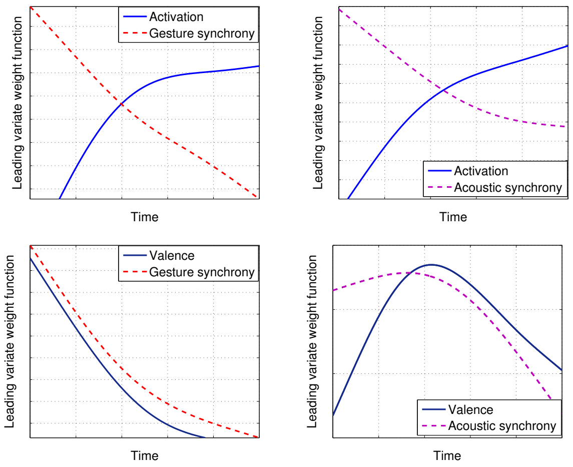

Figure 5: The smoothed leading canonical weight functions of emotion and behavior synchrony.

ture/acoustic synchrony curves, where each marker summarizes the overall dynamical pattern of the synchrony in an interaction. Different colors indicate distinct global turn-pair-wise emotion classes (see Section 2.2.2). Discernible dynamical synchrony patterns (esp. δ_2) are generally observed across emotion categories. For example, δ_2 of gesture synchrony in the low-valence dyads tends to locate in the upper part of the plane while that in the high-valence dyads generally lies in the lower part. We conduct t-tests to compare the PCA score distributions between emotion classes. Statistics indicate that the cross-emotion difference of synchrony dynamics is primarily expressed along the 2nd variation component — δ_2 distributions are significantly discriminative in the low and high activation/valence classes ($p < 0.05$) and there is no significant cross-emotion difference of δ_1 distributions. The distinguishability of δ_2 suggests that the dyadic emotions affect the temporal synchrony variability besides the overall synchrony strength and such emotion modulation is mainly reflected along the 2nd variation component.

4.2.3. Temporal Synchrony Dynamics in Relation to Continuous Dyadic Emotions

The above analysis presents a holistic picture about the interrelation between the temporal synchrony variability and the global dyadic emotions. Herein, we further study how such behavior synchrony and a dyad's emotions are dynamically correlated turn by turn over time.

FCCA is a useful tool for exploring the dynamical association between a pair of functional data [25] [24] [26]. It has also been applied to quantify the dynamical articulator-acoustic coupling in speech production [27]. In this work, we have two sets of functional data, synchrony curves $\{\widetilde{\sigma}_i(t)\}_i$ and emotional curves $\{\widetilde{e}_i(t)\}_i$. The objective of FCCA is to find a pair of functions $(\eta(t), \zeta(t))$ such that the canonical variates $(\int \eta(t)\widetilde{\sigma}_i(t), \int \zeta(t)\widetilde{e}_i(t))$ are maximally correlated,

$$\max_{(\eta(t),\zeta(t))} \frac{cov(\int \eta\widetilde{\sigma}_i, \int \zeta\widetilde{e}_i)^2}{[var(\int \eta\widetilde{\sigma}_i) + \lambda \int \mathcal{D}^2\eta][var(\int \zeta\widetilde{e}_i) + \lambda \int \mathcal{D}^2\zeta]}. \quad (3)$$

λ is a smoothing parameter. $\eta(t)$ and $\zeta(t)$ are the smoothed leading canonical weight functions, characterizing the temporal co-varying behavior of dyadic synchrony and emotion. The correlation between $\int \eta(t)\widetilde{\sigma}_i(t)$ and $\int \zeta(t)\widetilde{e}_i(t)$ is the smoothed leading canonical correlation r, measuring the strength of the emotion-synchrony association.

We apply FCCA to every two sets of emotion and synchrony data, *e.g.*, activation and acoustic synchrony. The smoothing parameter λ is selected in a cross-validation manner as introduced in [26]. Figure 5 shows the dynamical behavior of the smoothed leading emotion-synchrony weight functions. As can be seen, each pair of emotion-synchrony weights vary in a relatively similar manner over time, *i.e.,* there is a emotion-synchrony correlation at any particular time. For example, both valence-gesture synchrony weights decrease at a similar rate at any given time. This co-varying behavior of emotion-synchrony weights suggests that there generally exists a tight dynamical emotion-synchrony coupling over an interaction. Another interesting observation is that the valence-synchrony weights have an inphase relationship while the activation-synchrony weights co-vary in an antiphase way. This observation is congruent with the results in Table 1 that the low-activation dyads exhibit a tighter synchrony than the high-activation ones while the high-valence dyads show a stronger synchrony than the low-valence ones. In addition to the co-varying behavior exhibited in the emotion-synchrony weights, the smoothed shared variance between $\widetilde{e}(t)$ and $\widetilde{\sigma}(t)$ (*i.e.,* the squared smoothed canonical correlation r^2), corroborates a strong degree of temporal emotion-synchrony coupling: the valence-acoustic synchrony r^2 is .694; the valence-gesture synchrony r^2 is .719; the activation-acoustic synchrony r^2 is .697; and the activation-gesture synchrony r^2 is .645. These results imply that interactional synchrony is a prominent manifestation of the underlying affective processes, shedding further light toward informing affective interaction modeling and emotion recognition.

5. Conclusion

This work made an initial attempt at quantitatively investigating dynamic synchrony variability in gesture and vocal behavior in affective interactions. Our analysis results revealed that the interaction synchrony is a prominent manifestation of the underlying affective processes, from both holistic and turn-wise dynamic perspectives. The analysis brings us insight that can be useful for developing intelligent human-machine interaction, *e.g.,* creating an interaction computer that can generate appropriate behavior such that the human-machine behavior synchrony is controlled by their affective states. Also, in the future, we would like to incorporate synchrony dynamics into affect modeling for improving automatic emotion recognition.

6. References

[1] T. L. Chartrand and J. A. Bargh, "The chameleon effect: The perception–behavior link and social interaction." *Journal of personality and social psychology*, vol. 76, no. 6, p. 893, 1999.

[2] C. M. Murphy and T. J. O'Farrell, "Couple communication patterns of maritally aggressive and nonaggressive male alcoholics." *Journal of Studies on Alcohol*, vol. 58, no. 1, pp. 83–90, 1997.

[3] S. L. Johnson and T. Jacob, "Sequential interactions in the marital communication of depressed men and women." *Journal of Consulting and Clinical Psychology*, vol. 68, no. 1, p. 4, 2000.

[4] C.-C. Lee, A. Katsamanis, M. P. Black, B. R. Baucom, A. Christensen, P. G. Georgiou, and S. S. Narayanan, "Computing vocal entrainment: A signal-derived pca-based quantification scheme with application to affect analysis in married couple interactions," *Computer Speech & Language*, vol. 28, no. 2, pp. 518–539, 2014.

[5] Z. Yang, A. Metallinou, and S. Narayanan, "Analysis and predictive modeling of body language behavior in dyadic interactions from multimodal interlocutor cues," *Multimedia, IEEE Transactions on*, vol. 16, no. 6, pp. 1766–1778, 2014.

[6] R. Schmidt, S. Morr, P. Fitzpatrick, and M. J. Richardson, "Measuring the dynamics of interactional synchrony," *Journal of Nonverbal Behavior*, vol. 36, no. 4, pp. 263–279, 2012.

[7] R. Levitan and J. B. Hirschberg, "Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions," in *Interspeech*, 2011.

[8] A. Metallinou, Z. Yang, C.-C. Lee, C. Busso, S. Carnicke, and S. Narayanan, "The USC CreativeIT database of multimodal dyadic interactions: From speech and full body motion capture to continuous emotional annotations," *Language resources and evaluation*, 2015.

[9] Z. Yang and S. Narayanan, "Modeling dynamics of expressive body gestures in dyadic interactions," *Affective Computing, IEEE Transactions on*, 2016.

[10] I. Guide, "Autodesk®," 2008.

[11] S. Levine, C. Theobalt, and V. Koltun, "Real-time prosody-driven synthesis of body language," in *ACM Transactions on Graphics*, vol. 28, no. 5, 2009, p. 172.

[12] M. Sargin, Y. Yemez, E. Erzin, and A. Tekalp, "Analysis of head gesture and prosody patterns for prosody-driven head-gesture animation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 8, pp. 1330–1345, 2008.

[13] Z. Yang and S. Narayanan, "Modeling mutual influence of multimodal behavior in affective dyadic interactions," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, 2015, pp. 2234–2238.

[14] C. Busso, Z. Deng, S. Yildirim, M. Bulut, C.-M. Lee, A. Kazemzadeh, S. Lee, U. Neumann, and S. Narayanan, "Analysis of emotion recognition using facial expressions, speech and multimodal information," in *ACM International Conference on Multimodal Interaction*, 2004, pp. 205–211.

[15] R. Cowie, E. Douglas-Cowie, S. Savvidou*, E. McMahon, M. Sawey, and M. Schröder, "'feeltrace': An instrument for recording perceived emotion in real time," in *ISCA Tutorial and Research Workshop on Speech and Emotion*, 2000.

[16] A. Metallinou and S. Narayanan, "Annotation and processing of continuous emotional attributes: Challenges and opportunities," in *Automatic Face and Gesture Recognition (FG), IEEE International Conference and Workshops on*, 2013, pp. 1–8.

[17] W. M. Campbell, D. E. Sturim, and D. A. Reynolds, "Support vector machines using gmm supervectors for speaker verification," *Signal Processing Letters, IEEE*, vol. 13, no. 5, pp. 308–311, 2006.

[18] H. Li, G. Hua, Z. Lin, J. Brandt, and J. Yang, "Probabilistic elastic matching for pose variant face verification," in *Computer Vision and Pattern Recognition, IEEE Conference on*, 2013, pp. 3499–3506.

[19] T. S. Jaakkola, D. Haussler *et al.*, "Exploiting generative models in discriminative classifiers," *Advances in neural information processing systems*, pp. 487–493, 1999.

[20] F. Perronnin and C. Dance, "Fisher kernels on visual vocabularies for image categorization," in *Computer Vision and Pattern Recognition, IEEE Conference on*, 2007, pp. 1–8.

[21] P. J. Moreno and R. Rifkin, "Using the fisher kernel method for web audio classification," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, vol. 6, 2000, pp. 2417–2420.

[22] H. Kaya, A. A. Karpov, and A. A. Salah, "Fisher vectors with cascaded normalization for paralinguistic analysis," in *Interspeech*, 2015.

[23] 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, no. 3, pp. 115–126, 2006.

[24] J. O. Ramsay, *Functional data analysis*. Wiley Online Library, 2006.

[25] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3/4, pp. 321–377, 1936.

[26] S. E. Leurgans, R. A. Moyeed, and B. W. Silverman, "Canonical correlation analysis when the data are curves," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 725–740, 1993.

[27] Z. Yang, V. Ramanarayanan, D. Byrd, and S. Narayanan, "The effect of word frequency and lexical class on articulatory-acoustic coupling." in *Interspeech*, 2013, pp. 973–977.