HEAD MOTION SYNCHRONY AND ITS CORRELATION TO AFFECTIVITY IN DYADIC INTERACTIONS

Bo Xiao†, Panayiotis G. Georgiou†, Chi-Chun Lee†, Brian Baucom‡, Shrikanth S. Narayanan†

† SAIL, Dept. Electrical Engineering, University of Southern California, Los Angeles
‡ Dept. Psychology, University of Utah, Salt Lake City, U.S.A.
boxiao@usc.edu, georgiou@sipi.usc.edu, chiclee@usc.edu
brian.baucom@psych.utah.edu, shri@sipi.usc.edu

ABSTRACT

Behavioral synchrony, or entrainment, is a phenomenon of great interest to psychologists and a challenging construct to quantify. In this work we study the synchrony behavior of head motion in human dyadic interactions. We model head motion using Gaussian Mixture Model (GMM) of line spectral frequencies extracted from the motion vectors of the head. We quantify interlocutor head motion similarity through the Kullback-Leibler divergence of the GMM posteriors of their respective motion sequences. We use an audio-visual database of distressed couple interactions, extensively annotated by psychologists, to test two hypotheses using the derived similarity measure. We validate the first hypothesis — that people are more likely to increase their degree of synchrony as the interaction progresses — by comparing the first and second halves of the interaction. The second hypothesis tests if the relative change of the similarity measure from these two halves is significantly correlated with the behavioral annotation by the domain experts. This work underscores the importance of head motion as an interaction cue, and the feasibility of using it in a computational model for synchrony behavior.

Index Terms— Head motion; Synchrony; Behavioral signal processing; Entrainment; Linear prediction; Gaussian mixture model

1. INTRODUCTION

Many studies have reported on the phenomenon of unconscious synchrony amongst interlocutors in human interactions [1]. Interaction synchrony is also closely related to concepts such as entrainment, behavioral matching, mimicry, mirroring and so on. In general it means that certain behavioral aspects of the interlocutors become similar or coordinated during an interaction. Such behaviors are varied and multimodal, and include visual and vocal cues [2] of both verbal and non-verbal behavior [3]. Modeling synchrony is of interest in many applications, including notably psychotherapy [4], and interaction studies of mother-infant [5], teacher-student [6], group of musicians [7], *etc.* Theoretical descriptions of synchrony have largely been qualitative and abstract, and there is a great desire to devise a computational ancillary that can be empirically supported by behavior data. The present paper seeks to define such a measure for interaction synchrony using head motion cues.

Several studies have implicated synchrony as an underlying mechanism characterizing behavioral patterning such as positive affect and rapport in human interactions (*e.g.,* [2, 6, 8]). Particularly, in the scenario of psychotherapy, synchrony has also been connected with clinical outcomes [4]. These have motivated further detailed studies, and importantly, quantification of synchrony. The emerging field of Behavioral Signal Processing (BSP) [9] is focused towards studying human behaviors and in developing techniques to measure, analyze, and model human behavior signals to inform human assessment and decision making. It is within that framework, this work focuses on the quantitative study of synchrony in terms of deriving a signal measure that can explain specific behavioral patterns in human interactions.

A systematic study of behavioral synchrony needs to consider multimodal aspects of the interactions, feature design, as well as the choice of the appropriate computational approach. Delaherche *et al.* [10] have reviewed the multidisciplinary study of synchrony including computational approaches. Most previous studies reported focus on the design of automatic measurements of synchrony based on multimodal signal processing techniques, and validate these measurements through comparisons with human annotations or outcomes, or through comparisons with pseudo-interactions acting as control groups. The authors summarized mainly three types of methods to capture synchrony (entrainment, mimicry, *etc.*), such as computing the correlation of the signals from two interlocutors, comparing the phase and spectrum of the signals, and comparing bags of instances from each interlocutor. In addition, Lee *et al.* have proposed a vocal entrainment measure, which is based on the framework of finding the principal component analysis space of one interlocutor and projecting the features of the other one onto this space [2].

In this work, we study the synchrony of head motion using audio-visual data of real married couple interactions undergoing therapy. Head motion is a useful nonverbal behavioral cue present both when a person is a speaker and a listener in an interaction. There have been prior studies adopting head motion

This work is supported by NSF, DoD and NIH.

as a behavior cue for the analysis of synchrony. For instance Campbell [11] and Varni *et al.* [12] tracked the motion of the head as a time sequence signal. However, they did not adopt further segmentation or clustering of the motion.

Compared to previous works, we offer three contributions. First, we propose a data driven framework to establish a structure for characterizing head motion, via automatic clustering of motion types. Second, we define a quantitative measure for the degree of head motion similarity as a computational ancillary to interaction synchrony, which is based on the two bags of head motion instances from the two interlocutors. Third, we provide an empirical investigation of the proposed measure with natural dyadic human interaction data by testing the relative change of head motion similarity degree during the first and second halves of an interaction, and showing its correlation with psychologist annotated scores of behavioral codes where synchrony is inherently implicated as a moderating mechanism. We choose to analyze the relative change of similarity as a dynamic aspect because it is less studied, and not affected by the variability on the absolute values of similarity measure for different groups of interlocutors.

In the rest of the paper, we describe the data set used, the proposed head motion model and similarity measure, and the results of the experimental evaluation.

2. DATASET

The corpus used in this study comprises audio-visual recordings of seriously and chronically distressed couples having conversations about solving a problem in their marriage; it was collected during clinical studies conducted by the University of California, Los Angeles and the University of Washington [13]. Each couple talked about one problem chosen by the wife and another by the husband, for 10 minutes each. The analyzed data are from three points in time during the therapy process: before the psycho-therapy began, 26 weeks into the therapy and 2 years after the therapy session finished. The database includes 96 hours of recordings of 574 sessions. The video format is 704×480 pixels, 30 fps, with a screen split in the middle with one spouse on each side.

The behavior of both spouses in all sessions were characterized individually following two expert designed coding systems, the Couples Interaction Rating System 2 (CIRS2) [14] and the Social Support Interaction Rating System (SSIRS) [15]. The CIRS2 contains 13 behavioral codes and was specifically designed for conversations involving a problem in relationship, while the SSIRS consists of 20 codes that measure the emotional component of the interaction and the topic of conversation. The 33 codes are each on a numerical range from 1 to 9. Each session (and each code) was coded by at least three trained coders with a summative rating of the behavior of interest. In this paper we mainly focus on the "affectivity ratings" component of the SSIRS system since the role of synchrony is often implicated in affective behavior dynamics. We use the average score among coders as ground truth. Note that the codes only measure how much particular behavior patterns occur, independent of how much their opposite occur. For example, both *Global Positive* and *Global Negative* codes could have high value if they are both present in the interaction. A complete list of affectivity codes considered is shown in Tab. 1.

Table 1. Affectivity codes in Social Support Interaction Rating System (SSIRS) [15]

Global Positive, Global Negative, Anger/Frustration
Belligerence/Domineering, Contempt/Disgust, Sadness
Tension/Anxiety, Defensiveness, Affection, Satisfaction

The quality of the video recordings (done in different, real clinical settings) is not ideal; the relative positions of subjects as well as of the cameras are not fixed or known as the database was intended originally for human analysis. To mitigate data quality variability, we apply a preprocessing step to all sessions, separately on the left and right split screen content of the video. First, we run an OpenCV [16] face detector by uniformly sampling one frame per second from the video. Second, the face scale is estimated by the mode of the distribution of detected size of the face block. Third, we retain sessions that have a face detected on more than 70% of the sampled frames, and the estimated face scale is between 120 pixels and 160 pixels ($\frac{1}{4}$ to $\frac{1}{3}$ of image height). Estimation of head motion for sessions outside the above range of face scale is less reliable with the current approach. This resulted in 63 sessions (126 subjects) to be chosen. In these recordings, the upper body of a subject is captured while hands may or may not be in the field of view. One sample video frame is displayed in the top of Fig.1.

3. HEAD MOTION MODELING

3.1. Motion estimation

Face tracking and head motion estimation is a necessary front-end for later steps. As this module is not the focus of our work, we utilize a simple but effective setup. We first detect the face in each frame (marked by a square) using the cascade classifiers provided in OpenCV, and approximate face size with the side length of the square representing face. We slide a 5-frame window over the histogram of detected face sizes, and choose the face size $\hat{S}$ that maximizes the sum of the windowed histogram. In other words, we choose the most likely face size on a smoothed histogram. We exclude outliers of face detection by rejecting faces with size $S > 1.2\hat{S}$ or $S < 0.8\hat{S}$, which are very likely to be inaccurate in locating the head. The central location of face is estimated by the center of accepted faces. We again exclude faces with centers that are further than $\hat{S}$ on the horizontal axis or $0.5\hat{S}$ on the vertical axis to the estimated central location. Then we fill the gaps of frames missing a face by linear interpolation.

Optical flow of each pixel inside the face square is then computed, and finally the horizontal and vertical components of head motion are derived as the mean of horizontal and vertical optical flows over all pixels within that box, respectively. Given that the spouses remain in a sitting position throughout the session, this simple setup satisfies our needs and produces reliable results.

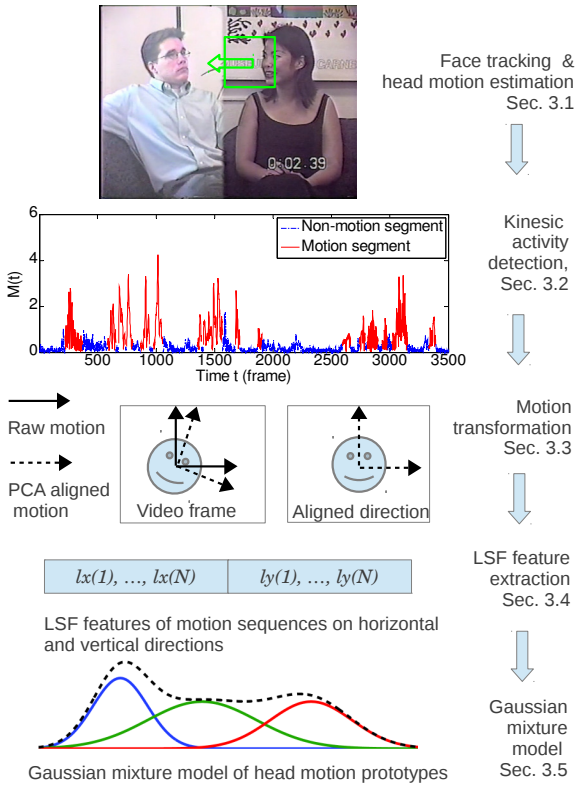

Fig. 1. Illustration of the processing steps in Sec.3

3.2. Kinesic activity detection

We set up a Kinesic Activity Detection (KAD) step to remove the segments where the spouse does not move. Let the horizontal and vertical components of the head motion stream be $M_x(t)$ and $M_y(t)$. We use the magnitude of motion $M(t) = \sqrt{M_x^2(t) + M_y^2(t)}$ as a 1D feature. We use a 2-Mixture GMM to represent motion versus non-motion classes, and a 2-state Hidden Markov Model (HMM) to represent the transition between the two classes. The parameters of the GMM are initialized by selecting the top 20% high valued $M(t)$ as being in the motion class while the rest are in the non-motion class; the initial transition probability of HMM is set to 0.9 for self-transition. The Expectation-Maximization (EM) algorithm is applied to obtain the maximum likelihood estimation of the states. Moreover, we post-process the state sequence by smoothing over short pauses (less than 0.2 seconds) when both sides of the detected pause are motion sequences longer than 1 second. Then we eliminate motion sequences that are less than 1 second, which are assumed to be noise in our model. An example of the KAD result is shown in Fig.1.

3.3. Motion transformation and windowing

Since the interacting spouses were sitting in arbitrary postures, the main directions of their head movements are hence not necessarily 0 or 90 degrees. We apply a Principal Component Analysis to the raw motion stream so as to align the main directions, as illustrated in Fig.1. We do a Z-normalization to the two dimensions of aligned motion streams (empirically found to be in a bell shape with heavy tails). The motion segments may have varying durations and contain a group of consecutive movements. These heterogeneous types of motions should be analyzed separately, so further segmentation within a motion segment is needed. Hence, for the analyses that follow, we apply a short time sliding window over each motion segment, with window length being 2 seconds, and window shift being 1 second. If the motion segment is less than 3 seconds then we do not window it further. Therefore, the windowed motion sequences could have length between 1 to 3 seconds.

3.4. Linear prediction features

We propose to use parameters derived through Linear Prediction (LP) as a transformed representation of the motion sequences for several reasons. First, assuming that head motion sequences can be viewed as being generated by an autoregressive process, LP offers a powerful way to capture the dynamic properties of various motion types. Second, LP is preferred over other methods such as vector quantization, because the motion sequences obtained through windowing may not be exactly aligned with the onset of motion. Third, LP provides the convenience of consistent feature dimension (equal to the LP order), while the windowed motion sequences are in varying lengths. We adopt the Line Spectral Frequencies (LSF) representation of LP [17], which is widely used in speech coding due to its better quantization properties.

We compute the LSF for horizontal and vertical components respectively, then concatenate the two components. In the couples interaction database we use for the study, the wives are denoted by $\mathtt{w}_1$ to $\mathtt{w}_{63}$, and husbands by $\mathtt{h}_1$ to $\mathtt{h}_{63}$. Let $j \in \{\mathtt{w}_1, \mathtt{h}_1, \mathtt{w}_2, \mathtt{h}_2, \cdots, \mathtt{w}_{63}, \mathtt{h}_{63}\}$ be a general index of any session and subject. Let $L_j^i = [Lx_j^i \; Ly_j^i]$ be the LSF of the i-th motion sequence in the j-th recording, and $Lx_j^i = \{lx_j^i(n)\}_{n=1}^N$ be the horizontal component, $Ly_j^i = \{ly_j^i(n)\}_{n=1}^N$ be the vertical component, where N is the order of LSF analysis. The constants $lx_j^i(0) \equiv 1$ and $ly_j^i(0) \equiv 1$ are omitted from the feature representation. As a result, we have a $2N$-dim feature for each motion sequence.

3.5. Gaussian mixture of head motion

Our goal is to compute a structure for characterizing head motion through statistical clustering. Here we construct a GMM with LSF features. We use the posterior probability of each feature instance as a soft cluster label. To train the GMM we pool all sessions together and conduct the training on all motion sequences. The K-mixture GMM is initialized by a K-means procedure, and iteratively optimized using the standard EM algorithm.

However, since the adopted GMM approach is unsupervised with the initialization being random, a single GMM does not guarantee good representation of head motion types. In the following sections, we base our analysis on multiple GMMs, which are initialized randomly on the same data.

Let π_k be the prior probability, $\mu_k = \{\mu_k(n)\}_{n=1}^{2N}$ be the mean vectors, and $\sigma_k = \{\sigma_k(n)\}_{n=1}^{2N}$ be the variance vector,

i.e., the diagonal of the assumed diagonal covariance matrix corresponding to the feature vector of dimension $2N$. The likelihood probability is given by $P(L_j^i|k) = \mathcal{N}(L_j^i; \mu_k, \sigma_k)$, and the posterior probability is given by Eq.(1).

$$P(k|L_j^i) = \frac{\pi_k P(L_j^i|k)}{\sum_{k'=1}^{K} \pi_{k'} P(L_j^i|k')}, k = 1 \cdots K. \quad (1)$$

4. HEAD MOTION SYNCHRONY MEASUREMENT

4.1. Similarity as lower tail of divergence distribution

For each subject in a session, there is a bag of motion sequences. Specifically, let $w \in \{\mathtt{w}_1, \mathtt{w}_2, \cdots, \mathtt{w}_{63}\}$, $h \in \{\mathtt{h}_1, \mathtt{h}_2, \cdots, \mathtt{h}_{63}\}$. We denote the bag of motion sequences of the wife w as $\mathcal{B}_w$, and that of the husband h as $\mathcal{B}_h$, *e.g.,* $\mathcal{B}_{\mathtt{w}_1} = \{L_{\mathtt{w}_1}^i\}_{i=1}^I$, where I is the total number of motion sequences for $\mathtt{w}_1$.

We adopt the Kullback-Leibler (KL) divergence to compute the similarity between two motion instances L_w^i and $L_h^{i'}$ as in Eq.(2). In order to avoid numerical instability caused by zero values in the posterior, we add a small positive value $\epsilon = 1 \times 10^{-5}$ to all elements of the posterior probability and re-normalize.

$$\mathtt{KL}(L_w^i, L_h^{i'}) = \sum_{k=1}^{K} P(k|L_w^i) \mathtt{log} \frac{P(k|L_w^i)}{P(k|L_h^{i'})} \quad (2)$$

Based on the KL divergence, we obtain the proposed similarity measure. Let the similarity function over the two bags of motion sequences be $\mathtt{sim}(\mathcal{B}_w, \mathcal{B}_h, \rho)$, where ρ is a percentile parameter. The procedure is described as follows.

1. Compute pairwise KL divergence for all pairs of motion instances in $\mathcal{B}_w$ and $\mathcal{B}_h$, resulting in a matrix $SIM_{I \times I'} = \{\mathtt{KL}(L_w^i, L_h^{i'})\}$.
2. Convert the matrix $SIM_{I \times I'}$ to a single vector and sort by ascending order, resulting in a new vector S.
3. Obtain $\mathtt{sim}(\mathcal{B}_w, \mathcal{B}_h, \rho)$ as the mean value of S from the smallest element to the ρ percentile.

The motivation of having the parameter ρ is to capture any similar motion sequences, regardless of if there are very different ones in the same bags of motion sequences. For computational simplicity we do not match motion sequences one-to-one or group-to-group, but take the mean of pairwise KL divergence as an averaged measure. We also obtain $\mathtt{sim}(\mathcal{B}_h, \mathcal{B}_w, \rho)$ in a similar manner. Note that it is not equivalent to $\mathtt{sim}(\mathcal{B}_w, \mathcal{B}_h, \rho)$ as KL divergence is non-symmetric.

4.2. Dynamics of synchrony degree

We analyze the synchrony dynamics during the interaction at a very basic level, by computing the relative change of similarity measure over the first and second halves of the session, denoted by R. We do not further divide the session in this study, since the motion instances may become sparse in short duration. The bags of motion sequences in the first and second halves are denoted $\mathcal{B}_w^1$ and $\mathcal{B}_w^2$ for the wife, respectively, and similarly for the husband.

Due to the unsupervised nature of the derived GMM, a single GMM derived by random initialization might not converge in the direction aligned with the behaviors of interest. Therefore, we train an ensemble of M GMMs over the same data with different initializations, and use the average of R as a more robust measure. Let a single $\tilde{R}$ be the result of one GMM, as in Eq.(3). We then define $R(w, h)$ as in Eq.(4), converted to a log scale. Similarly we can obtain $R(h, w)$. Note that a smaller value of $R(w, h)$ or $R(h, w)$ means lower divergence and greater similarity.

$$\tilde{R} = \frac{\mathtt{sim}(\mathcal{B}_w^2, \mathcal{B}_h^2, \rho)}{\mathtt{sim}(\mathcal{B}_w^1, \mathcal{B}_h^1, \rho)} \quad (3)$$

$$R(w, h) = \log(\frac{1}{M} \sum_{m=1}^{M} \tilde{R}_m) \quad (4)$$

5. HYPOTHESES TESTS AND EXPERIMENT RESULTS

To explore the usefulness of the proposed similarity measure, we consider two hypotheses, inspired by qualitative descriptions offered by domain experts. The first one tests the hypothesis that synchrony increases as the interaction progresses. The second one tests the relation between the relative change of the proposed measure and affective behavior, where synchrony has been theoretically implicated by psychologists.

5.1. Hypotheses tests

We consider the relative change of similarity measure on an individual basis. Let $\mathcal{R} = \{R(\mathtt{w}_l, \mathtt{h}_l)\} \cup \{R(\mathtt{h}_l, \mathtt{w}_l)\}$ where $l = 1, 2, \cdots, 63$. Let $\mathcal{Y} = \{Y(\mathtt{w}_l)\} \cup \{Y(\mathtt{h}_l)\}$ be the behavior codes on affectivity as in Tab. 1. The KL divergence $\mathtt{KL}(L_h^i, L_w^{i'})$ measures the information loss when the posterior of $L_w^{i'}$ is used to approximate that of L_h^i [18]. In other words, when the husband's movements are considered as reference, we use the wife's movements to approximate those. If the wife has a high level of synchrony to the husband, then the approximation should come with low divergence, hence we investigate the relation of the $R(\mathtt{h}_l, \mathtt{w}_l)$ value with the wife's behavior codes $Y(\mathtt{w}_l)$. Similarly, $Y(\mathtt{h}_l)$ is viewed in relationship to $R(\mathtt{w}_l, \mathtt{h}_l)$.

The first hypothesis test is binomial test on whether $r \in \mathcal{R}$ is more likely to be smaller than 0. The two hypotheses are:

H1$_0$ whether $r \geq 0$ or $r < 0$ is random, *i.e.,* 50% chance.

H1$_a$ there are more subjects having $r < 0$, *i.e.,* as interactions progress it is more likely that divergence decreases, hence synchrony increases.

The second hypothesis test uses the Student's t-distribution to test the presence of Pearson's correlation coefficients between $\mathcal{R}$ and $\mathcal{Y}$. The two hypotheses are:

H2$_0$ $\mathcal{R}$ and $\mathcal{Y}$ are uncorrelated.

H2$_a$ there is some correlation between $\mathcal{R}$ and $\mathcal{Y}$.

5.2. Experiment setting

The order of LSF in our feature analysis was set to 10, as we empirically found higher order does not offer any further advantage. We set the GMM ensemble size M to 50, as a trade-off between having an adequate number of GMMs to increase robustness and assuring affordable computation.

Previous research in psychology has suggested *six* classes of head motion [19]. Guided by this, we employed 4 to 12 mixtures for each GMM ($K = 4, 5, \cdots, 12$), since we expected the automated clustering to capture a finer structure of head motion than manual labeling. We conducted the experiments with a sampling on the parameter ρ ($0 < \rho < 1$) from 0.05 to 0.5, with a step size 0.05.

We found that there were sessions with extreme values of R, which were found to have very noisy video data. In order to avoid the influence of outliers in computing correlation, we exclude sessions of the top 3% largest $|r|$ values (on two tails in logarithm domain) for each test of correlation.

Moreover, to verify that any found correlations are meaningful effects of the interaction, we conducted the correlation analysis with random pairings of wives and husbands. In other words, the synchrony degree and relative change were also computed based on subjects who did not meet with each other (*i.e.,* not "true" couples). We repeated the shuffling for 100 times.

5.3. Results and discussion

We compute the percentile of subjects having $r < 0$ as in hypothesis $\mathbf{H1}$, summarized in Fig. 2. Each row represents one value of K and each column represents one value of ρ. In binomial test with sample size 122 (excluding outliers), percentile of 61% has one-tail p-value of 0.01. Therefore, $\mathbf{H1}_0$ is rejected with $p \leq 0.01$ for all parameter choices but for a few exceptions. This supports the notion that the phenomenon of head motion synchrony in dyadic interaction is reflected by a tendency of increased similarity of head motion towards the other interlocutor.

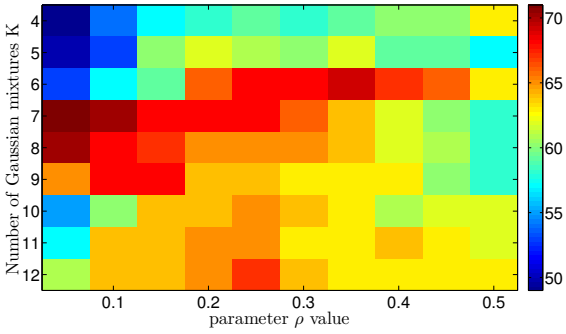

Fig. 2. Percentile of subjects having $r < 0$

We next compute Pearson's correlation of the proposed synchrony measure $\mathcal{R}$ and the expert-specified behavior code $\mathcal{Y}$ under various experimental settings. The correlations are more consistently significant with $\rho = 0.05$. Among the affectivity codes, *Global Positive, Global Negative, Affection, Satisfaction* are significantly correlated with $\mathcal{R}$, in addition to the code of *Use of humor*. The correlation obtained with varying K and ρ are shown in Fig. 3.

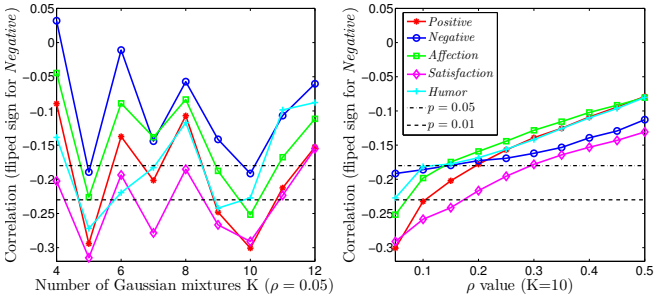

Fig. 3. Correlation between $\mathcal{R}$ and $\mathcal{Y}$

Table 2. Correlation of $\mathcal{R}$ and $\mathcal{Y}$ in random pairing of interlocutors ($K = 10$, $\rho = 0.05$)

Code	*Posit.*	*Negat.*	*Affect.*	*Satisf.*	*Humor*
Mean	-0.01	-0.02	0.02	0.01	-0.01
Std	0.09	0.08	0.09	0.10	0.08

Specifically, positive affect codes are negatively correlated with R, *i.e.,* the more positive attitude the subjects shows, the smaller the derived divergence measure R is. In other words, spouses having positive affect are associated with an increasing degree of synchrony along the interaction. Similarly, the negative affect code *Global Negative* is positively correlated with $\mathcal{R}$, suggesting that spouses having negative affect are associated with a decreasing degree of synchrony along the interaction.

The higher significance with $\rho = 0.05$, *i.e.,* only the 5% smallest divergence pairs, lends support to the intuition that capturing the most salient synchrony pairs of motion is more important than taking the overall average of pairwise divergence. We also see in this experiment that the correlations are in general high with $K = 5$ and $K = 10$. We want to further study the effect of K in future.

In addition, we report the experiment results of random pairing with $K = 10$ and $\rho = 0.05$. In Tab. 2 the mean and standard deviation of correlations are listed for each affectivity code. None of the mean correlations are significant ($p < 0.05$), and any significant value is beyond at least one standard deviation. Furthermore, there are 83, 14, 2, 0 and 1 times that random pairings have, respectively 0, 1, 2, 3 and 4 codes associated with significant correlation. Compared to the result in Fig. 3, the results of the randomized experiment suggest that although a spurious high correlation could exist in some randomly paired interactions, this does not generally happen.

6. CONCLUSION

We studied the synchrony phenomenon of head motion manifested in human dyadic interactions using a quantitative measure of similarity. To facilitate the analysis, we designed a structural model of head motion, using Gaussian Mixture

Models over line spectral frequencies of head motion sequences. Based on the GMMs, any two head motion instances are compared through KL divergence of their respective model posterior probabilities. We define a similarity measure of head motion utilizing the KL divergence. Finally, we compute the relative change of head motion similarity during the first half and the second half of the interaction. Experiment results focused on two aspects of the proposed synchrony measure in light of qualitative descriptions that have been offered by behavioral science experts. First, there are more subjects with a higher degree of synchrony (a smaller value of divergence) in the second half of the conversation, which suggests a rise of synchrony as the interaction progresses, consistent with what is expected. Second that the relative change in the similarity measure is correlated with multiple behaviors as annotated by human experts, and where synchrony dynamics have been implicated in describing the underlying behavior mechanisms. These findings demonstrate the usefulness of the proposed head motion modeling approach and the derived synchrony measure.

The study of head motion as an indicator of specific behavior expression is a topic of on-going work, and much remains to be done. For example, in our future work we would like to analyze the derived GMM classes in collaboration with domain experts to better interpret the automatic modeling results. We would also like to study the dynamics of synchrony over finer temporal resolutions and investigate the relation of synchrony with finer behavior traits such as reactivity. Moreover, we have so far studied synchrony in acoustic and visual modalities separately, but we intend to investigate joint models of multimodal behaviors in the future.

7. REFERENCES

[1] T.L. Chartrand and R. van Baaren, "Human mimicry," *Advances in experimental social psychology*, vol. 41, pp. 219–274, 2009.

[2] 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*, 2012.

[3] X. Sun, K.P. Truong, M. Pantic, and A. Nijholt, "Towards visual and vocal mimicry recognition in human-human interactions," in *Proc. SMC*. IEEE, 2011, pp. 367–373.

[4] F. Ramseyer and W. Tschacher, "Nonverbal synchrony in psychotherapy: coordinated body movement reflects relationship quality and outcome," *Journal of consulting and clinical psychology*, vol. 79, no. 3, pp. 284, 2011.

[5] F.J. Bernieri, J.S. Reznick, and R. Rosenthal, "Synchrony, pseudosynchrony, and dissynchrony: Measuring the entrainment process in mother-infant interactions," *Journal of Personality and Social Psychology*, vol. 54, no. 2, pp. 243, 1988.

[6] F.J. Bernieri, "Coordinated movement and rapport in teacher-student interactions," *Journal of Nonverbal behavior*, vol. 12, no. 2, pp. 120–138, 1988.

[7] A. Camurri, G. Varni, and G. Volpe, "Measuring entrainment in small groups of musicians," in *Proc. ACII*. IEEE, 2009, pp. 1–4.

[8] M. LaFrance, "Nonverbal synchrony and rapport: Analysis by the cross-lag panel technique," *Social Psychology Quarterly*, pp. 66–70, 1979.

[9] S. Narayanan and P. Georgiou, "Behavioral signal processing: Deriving human behavioral informatics from speech and language," *Proceedings of the IEEE*, vol. 101, no. 5, pp. 1203–1233, May 2013.

[10] E. Delaherche, M. Chetouani, A. Mahdhaoui, C. Saint-Georges, S. Viaux, and D. Cohen, "Interpersonal synchrony: A survey of evaluation methods across disciplines," *to appear in IEEE Transactions on Affective Computing*, 2012.

[11] N. Campbell, "An audio-visual approach to measuring discourse synchrony in multimodal conversation data," in *Proc. Interspeech*, 2009, pp. 2159–2162.

[12] G. Varni, G. Volpe, and A. Camurri, "A system for real-time multimodal analysis of nonverbal affective social interaction in user-centric media," *IEEE Transactions on Multimedia*, vol. 12, no. 6, pp. 576–590, 2010.

[13] A. Christensen, D.C. Atkins, S. Berns, J. Wheeler, D.H. Baucom, and L.E. Simpson, "Traditional versus integrative behavioral couple therapy for significantly and chronically distressed married couples," *Journal of consulting and clinical psychology*, vol. 72, no. 2, pp. 176–191, 2004.

[14] C. Heavey, D. Gill, and A. Christensen, "Couples interaction rating system 2 (cirs2)," *University of California, Los Angeles*, 2002.

[15] J. Jones and A. Christensen, "Couples interaction study: Social support interaction rating system," *University of California, Los Angeles*, 1998.

[16] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[17] P. Kabal and R.P. Ramachandran, "The computation of line spectral frequencies using chebyshev polynomials," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 34, no. 6, pp. 1419–1426, 1986.

[18] K.P. Burnham and D.R. Anderson, *Model selection and multi-model inference: a practical information-theoretic approach, 2nd Ed.*, pp. 50–54, Springer, 2002.

[19] J.A. Harrigan, R. Rosenthal, and K.R. Scherer, *The new handbook of Methods in Nonverbal Behavior Research*, pp. 137–198, Oxford University Press, New York, 2005.