Modeling Head Motion Entrainment for Prediction of Couples' Behavioral Characteristics

Bo Xiao*, Panayiotis Georgiou*, Brian Baucom† and Shrikanth Narayanan*
*SAIL, Dept. Electrical Engineering, University of Southern California, Los Angeles, CA 90089
†Dept. Psychology, University of Utah, Salt Lake City, UT 84112, U.S.A.
boxiao@usc.edu, georgiou@sipi.usc.edu, brian.baucom@psych.utah.edu, shri@sipi.usc.edu

***Abstract*—Our work examines the link between head motion entrainment of interacting couples and human expert's judgment on certain overall behavioral characteristics (*e.g., Blame* patterns). We employ a data-driven model that clusters head motion in an unsupervised manner into elementary types called kinemes. We propose three groups of similarity measures based on Kullback-Leibler divergence to model entrainment. We find that the divergence of the (joint) distribution of kinemes yields consistent and significant correlation with target behavior characteristics. The divergence of the conditional distribution of kinemes is shown to predict the polarity of the behavioral characteristics. We partly explain the strong correlations via associating the conditional distributions with the prominent behavioral implications of their respective associated kinemes. These results show the possibility of inferring human behavioral characteristics through the modeling of dyadic head motion entrainment.**

***Keywords*—Head motion; Entrainment; Kineme; Similarity; Behavioral characteristics**

I. Introduction

Entrainment, also known as synchrony, is the ability of a human to "detect a mind and resonate with its outputs" nonconsciously [1]. This means that the behavioral expressions, *e.g.,* the spoken and gestural communicative signals of one interlocutor, can be perceived by the other interlocutor and cause them to resonate their outputs, *e.g.,* by employing similar words, spoken patterns or gestures [2]. Behavioral entrainment entails, and fosters, neural entrainment [1]. Studies on entrainment from three perspectives — cognitive models of imitation, social psychological studies on mimicry, and empirical findings from neurosciences — are converging to a consistent theory of shared representational format of perception and action. Researchers believe such a mechanism has adaptive advantages of "understanding the other mind", which facilitates intersubjectivity and social behavior [3].

In practice, analyzing entrainment is also of interest in many application domains, including notably psychotherapy [4], studies of interactions amongst mother-infant [5], teacher-student [6], and even a group of musicians [7]. In psychotherapy studies, behavioral entrainment serves as an indication of rapport between interlocutors and hence relates to the outcome of therapy. For example, in [8], satisfied couples showed stronger associations (compared to dissatisfied ones) between the two partners' respective change in levels of immediacy as measured through gaze direction, body openness, and body position (*e.g.,* looking at partner, open body, leaning towards partner corresponded to high immediacy behaviors). Ramseyer and Tschacher [4] also showed that motion synchrony was correlated with the quality of a relationship and reduction of negative outcomes.

Computationally modeling and quantifying entrainment in an objective manner can be a valuable tool for psychotherapy, given its importance as a behavioral cue. Further, machine processing can observe cues from multiple streams and at speed much higher than human capabilities; and can enable scalability, potentially offering novel insights to the human expert. To achieve this goal, there are several technical challenges to address, including the representation and modeling of behavior, the measurement of behavioral aspects of signal similarity while discarding other aspects of the signal, the interplay of multimodal cues, and the evaluation of the quality of the derived metric. Entrainment is one of a larger set of behavioral mechanisms that can be investigated under the framework of *Behavioral Signal Processing* (BSP) as introduced by Narayanan and Georgiou [9]. In short BSP refers to "techniques and computational methods that support the measurement, analysis, and modeling of human behavior signals that are manifested in both overt and covert multimodal cues (expressions), and that are processed and used by humans explicitly or implicitly (judgments and experiences)". Our work is posed within the BSP framework and examines one aspect of entrainment behavior, as manifested in head motion of the interlocutors.

Quantification of entrainment has been studied by many researchers. Sun *et al.* studied mimicry in face-to-face conversations, shedding light on the mimicry of visual and non-verbal vocal behaviors [10]. Delaherche *et al.* [11] have summarized a range of methods used for capturing entrainment. They categorized those into three types based on the comparison of the signals from the two interlocutors: correlation based, phase and spectrum comparison, and bags-of-instances comparison. In addition, Lee *et al.* [12] have proposed a vocal similarity measure of interlocutors, which employed Kullback-Leibler Divergence (KLD) of energy distribution in the Principal Component Analysis spaces derived by acoustic features. The method is particularly amenable to handling the possibly asynchronous and discontinuous nature of the cues exchanged between the interlocutors.

Among visual cues of human behaviors, facial expressions and gaze are well studied and counted for communicating emotions [13], [14]. Nevertheless, head motion is also an

This work is supported by NIH, NSF and DoD.

important behavioral cue that is less studied [15], [16]. Head motion conveys rich implicit and some explicit communicative functions [17]. In order to transcribe continuous motion into discrete units, Birdwhistell [18], a psychologist, proposed the theory of "kinesic-phonetic" analogy, which defined *kinemes* as the elementary units of motion, similar to phonemes (the elements of language's phonology such as vowels and consonants). For example, one can assign a kineme to represent a head sweep to the right. According to Birdwhistell, consecutive kinemes can be combined to form *kine-morphs*, or even larger units of *kinemorphic constructions*; thus nonverbal behavior can be treated in the same way as verbal language.

However, it has proven difficult to find a commonly agreed inventory of head motion types that is comprehensive and convenient to use [19]. From an engineering point of view, we have proposed a data-driven clustering method for constructing a structure of head motion types to address this issue [20]. Our approach applied shifting windows to the head motion signal, represented the signal with power-spectral features, and clustered the motion types using K-means algorithm. Distribution of the motion events in the clusters by a subject was able to predict the subject's behavioral characteristics, showing that the head motion model is effective in capturing behavioral cues. In [21] we have also proposed a similarity measure between two motion events based on KLD, which was averaged to obtain the overall similarity measure in an interaction session. Further we showed a link between the behavioral characteristics and the relative change of similarity levels in the two halves of an interaction.

In this work we employ the head motion clustering model introduced in [20] to represent the motion signals. Our main contributions are as follows. First, we propose three groups of features for computing KLD based similarity measures: the distribution of kinemes for each interlocutor, the joint distribution of neighboring kinemes in time, and the conditional distribution of kinemes given the previous ones. The later two groups try to capture the temporal dynamics of head motion within a single interlocutor or between two interlocutors. Second, we investigate the following questions: (1) are these similarity measures correlated with certain target behavioral characteristics? and (2) can these similarity measures predict the expert-specified behavioral characteristics? We employ an expert-annotated database of real couples' conversations in couple therapy in the experiments. The overview of the system flow is shown in Fig. 1.

In the rest of the paper, we briefly describe the employed dataset in Sec. II, and the head motion model in Sec. III. We propose the similarity measures in Sec. IV. We report the experiment design and results in Sec. V. Furthermore, we discuss about the saliency of kinemes in Sec. VI. We conclude the paper with remarks of future directions in Sec. VII.

II. Dataset

Psychologists from the University of California, Los Angeles and the University of Washington conducted a longitudinal research study of couples therapy [22]. They recorded the conversations of 134 chronically distressed couples discussing problems in their marriage, where the wife and the husband chose a topic to discuss in turn, for 10 min each. A subsample

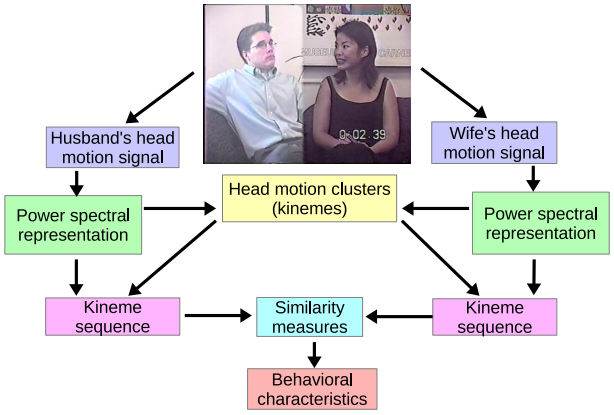

Fig. 1: Overview of the system flow

of 117 couples is available for the current study. The audio-visual recordings were collected at three time stages during the study: before the therapy, 26 weeks into the therapy, and 2 years afterwards. The entire database contains 96 hours of recording in 574 sessions. The video format was DV-NTSC, 704×480 pixels, 30 frames-per-second, with a screen split and one spouse on each side taking a sitting posture.

At least three psychology undergraduate students viewed each session in the database, and coded several aspects of each individual interlocutor's behavioral characteristics on a discrete scale from 1 to 9. These students were well trained for the task, and were considered domain experts. The behavior codes were defined in two coding manuals: the Couples Interaction Rating System 2 (CIRS2) [23] which is specifically designed for conversations involving a problem in relationship; and the Social Support Interaction Rating System (SSIRS) [24] which measures the emotional component of the interaction and the topic of conversation. The codes were designed at interaction session level, offering *overall* assessments of target behavioral characteristics. Local (within interaction) level annotation is not available from the database. In this paper we select 4 codes which associate with behavioral entrainment as pointed out by theoretical analyses [25]–[27]: *Acceptance*, *Blame* from CIRS2, and *Positive*, *Negative* from SSIRS [1]. We use the average score among coders as ground truth. Note that assessments of the codes were independent, although the code values may correlate.

Our focus is on the visual behavioral modality. The database was collected in the 1990s from various clinical settings, originally intended for manual analysis. Therefore the video quality is not ideal for automated processing. Furthermore there was no calibration and the relative positions of subjects and the cameras were not available. Thus for the experiments in this work, we employ 221 sessions (about 37 hours) from the database, which have acceptable quality of head tracking for both spouses.

[1] *Acceptance* indicates understanding and acceptance of partner's views, feelings and behaviors. *Blame* indicates that one blames, accuses, or criticizes the partner, uses critical sarcasm, makes character assassinations. *Positive* and *Negative* are overall rating of the positive and negative affect the target spouse showed during the interaction. Examples include overt expressions of warmth, support, acceptance, affection, positive negotiation and compromise for *Positive*, and rejection, defensiveness, blaming, and anger for *Negative*.

III. Head motion model

A. Head motion estimation

Head motion estimation involves "corner-like" point tracking and pruning of such tracks according to detected face regions, implemented in OpenCV [28]. First, we repeatedly search for "good-features-to-track" [29] at a fixed interval of every 5 frames. We estimate the optical flow [30] of the above feature points in neighboring frames. For any valid temporal trajectory, we require the forward-time and backward-time optical flows to overlap, *i.e.,* the backward optical flow that is starting from the pixel predicted by the forward optical flow should end back to the original feature point.

Second, we detect faces in each frame using Haar cascade classifiers. Let τ be the number of frames having a detected face during the life span of a track. We keep any track that resides in the face region in more than 0.9τ frames. We remove long-duration, stationary tracks that are likely to be in the background. This approach yields head motion tracks covering 97% of frames in average in time.

Third, the motion signal is estimated by averaging the motion of all tracks in each frame. We normalize the motion signal by the average side length of the detected square face regions, to compensate for the variability of the subject-to-camera distances. We denote the head motion signal on the horizontal (X-dir) and vertical (Y-dir) directions as $M_x(n)$ and $M_y(n)$, respectively.

B. Power spectral analysis

Hadar *et al.* [31] suggested that magnitude and frequency are plausible dimensions to quantify head motion. To capture these cues, we compute the power spectrum of the head motion signal, which represents the energy distribution of the signal against frequency.

We apply a window of 2 seconds (60 frames) with 1 second (30 frames) overlap on $M_x(n)$. For each window, we obtain an autocorrelation function $R(M_x, l)$ for $M_x(n)$ of length 119. We compute the power spectrum $S(M_x, f)$ as the absolute value of 128-point Discrete Fourier Transform of $R(M_x, l)$. The first 64 points of $S(M_x, f)$, *i.e.,* $f \in \{0, 1, \cdots, 63\}$, correspond to the frequency band of 0 to 15 Hz. We are only interested in the $S(M_x, f)$, $f \in [1, 15]$ band that corresponds to movement below 3.5 Hz, the rest being either stationary or of head movement speed unlikely to be observed.

In the same way, we obtain the power spectrum for Y-dir $S(M_y, f)$, $f \in [1, 15]$, for each sliding window. Finally, we convert the power spectrum to log scale.

C. Head motion clustering

We construct the head motion model by clustering all the motion events in the training data. Let $\mathcal{M}$ be the set of motion events represented by $S(M_x, f)$ and $S(M_y, f)$. We compute the mean μ, variance σ^2, and zero-mean, unit-variance normalization $\overline{\mathcal{M}}$, for X-dir and Y-dir separately. We iteratively set the number of clusters K from 4 to 25, *i.e.,* from clustering at a coarse level to clustering at a fine level — finer than commonly adopted by psychologists in coding head motion.

Given K, we initialize the cluster centroids $\{C_i\}_{i=1}^{K}$ on a randomly selected 10% of the training data $\overline{\mathcal{M}}$, then optimize on the entire $\overline{\mathcal{M}}$, repeating for 5 times in order to find the best clusters with minimum total distance (Euclidean) between all samples and their associated cluster centroids. According to [20], we cluster head motion events of wives and husbands together, but for X-dir and Y-dir separately. Such a scheme allows representation of the wife and the husband's motions with a common set of kinemes, while capturing the difference of motion types in horizontal and vertical directions.

IV. Entrainment model

It is widely accepted that entrainment can be approximated by measuring behavioral similarity; however, the design of an appropriate similarity measure is still an open research question. In this section we propose three groups of similarity measures based on the aforementioned head motion model.

For an analysis window with index t, we first identify a kineme by finding the nearest cluster centroid using the distance in the power spectral feature domain. We then represent the X-dir kineme sequence $w_x(t)$ for the wife in one dyad as in (1), where T is the total number of analysis windows for the dyad. Similarly, $w_y(t)$ is the Y-dir kineme sequence for the wife; and $h_x(t)$, $h_y(t)$ are the corresponding ones for the husband.

$$w_x(t) = \arg\min_i |C_i - S(M_x(t), f)| \qquad (1)$$
$$i = 1, 2, \cdots, K; \quad t = 1, 2, \cdots, T.$$

The experiment is evaluated for each assumed number of kineme types, K, and we denote that as $\{w_x(t)\}^K$, $\forall K \in \mathbf{N}^+$, $4 \leq K \leq 25$. We omit $\{\cdots\}^K$ for simplicity when the same operation is applied for every value of K. We consider a symmetric similarity measure for both the wife and the husband, formulated as symmetric Kullback-Leibler Divergence (KLD) in function $\texttt{Div}(\cdot, \cdot)$ as in (2) for the example of two-variable case.

$$\texttt{Div}(P, Q) = \frac{1}{2}\sum_{x,y} P(x, y) \log \frac{P(x, y)}{Q(x, y)} \quad + \quad \frac{1}{2}\sum_{x,y} Q(x, y) \log \frac{Q(x, y)}{P(x, y)} \qquad (2)$$

Here P and Q are random variables representing a *conjugate pair* of behavior descriptors for the wife and the husband, *e.g.,* w_x and h_x. We list the paired behavior descriptors in Table I and Fig. 2.

A. Distribution of kinemes

There are two conjugate pairs of behavior descriptors corresponding to the marginal distribution of kinemes, *i.e.,* w_x *vs.* h_x, and w_y *vs.* h_y. We define $P(w_x)$ as in (3).

$$P(w_x = i) = \frac{1}{T+1}\Big(\frac{1}{K} + \sum_{t=1}^{T} I(w_x(t), i)\Big) \qquad (3)$$
$$i = 1, 2, \cdots, K.$$

TABLE I: Conjugate pairs of behavior descriptors, $(u, v) \in \{(x, x), (x, y), (y, x), (y, y)\}$, 10 and 8 pairs in total for joint and conditional distribution, respectively, for the wife and husband dyad.

Joint distribution			Conditional distribution		
Conjugate pairs		#	Conjugate pairs		#
(w_x, w_y)	(h_x, h_y)	1	-	-	-
(w_x, h_y)	(h_x, w_y)	1	-	-	-
$(w_u^{'}, w_v)$	$(h_u^{'}, h_v)$	4	$w_v \vert w_u^{'}$	$h_v \vert h_u^{'}$	4
$(w_u^{'}, h_v)$	$(h_u^{'}, w_v)$	4	$h_v \vert w_u^{'}$	$w_v \vert h_u^{'}$	4
Total count		10	Total count		8

$I(\cdot, \cdot)$ denotes identity function, which takes value 1 when the two arguments are identical, otherwise 0. To avoid zero probabilities, we add $\frac{1}{K}$ uniformly to all entries of the counts. This can be viewed as a smoothed count-and-divide approach. We obtain the divergence between w_x and h_x as in (4), similarly for $D_d(w_y||h_y)$.

$$D_d(w_x||h_x) = \texttt{Div}(P(w_x), P(h_x)) \tag{4}$$

B. Joint distribution of two associated kinemes

Further to the marginal distribution, we consider kineme pairs within or across the interlocutors that may have dependencies, as summarized in Table I (left side). The notion of behavior descriptors and conjugate pairs are illustrated in Fig. 2 (the kinemes with prime symbols are in the previous time point). The first two conjugate pairs in Table I (left side) are at the same time point, while the rest are across two consecutive time points. We estimate the joint probability, *e.g.,* $P(w_x, w_y)$ in (5), based on the occurrence of the pair in the kineme sequences $w_x(t)$ and $w_y(t)$.

$$P(w_x = i, w_y = j) = \frac{1}{T+K}\Big(\frac{1}{K} + \sum_{t=1}^{T} I(w_x(t), i) I(w_y(t), j)\Big) \tag{5}$$
$$i = 1, 2, \cdots, K; \quad j = 1, 2, \cdots, K.$$

We obtain 10 divergences for each K, 220 in total for $K = 4, 5, \cdots, 25$, computed as for instance in (6).

$$D_j\big((w_x, w_y)\big|\big|(h_x, h_y)\big) = \texttt{Div}\big(P(w_x, w_y), P(h_x, h_y)\big) \tag{6}$$

C. Distribution of kinemes conditioned on history

The above joint distribution based diversity D_j treats every kineme equally. The behavioral similarity exhibited in reaction to some kinemes may be more important than others. Therefore, we consider the distribution of the current kineme (*e.g.,* w_x) conditioned on the one in the previous time point (*e.g.,* $w_x^{'}$, with the prime symbol denoting the previous time point), as shown in Table I (right side) and Fig. 2. We estimate

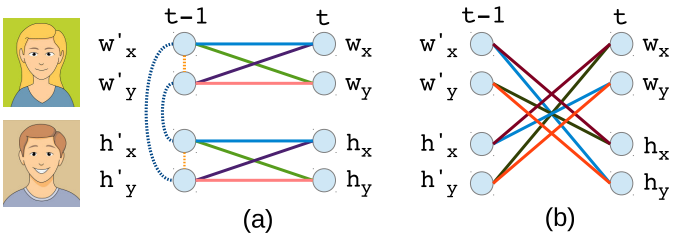

Fig. 2: Illustration of conjugate pairs of behavior descriptors identified by the same color. (a) two conjugate pairs at the same time point, and four pairs of "self-context", the temporal relation of the subject's behavior. (b) four pairs of "cross-context", the reaction to the partner's stimuli.

the conditional probability, *e.g.,* $P(w_x|w_x^{'})$ in (7), based on the occurrence of kinemes following a certain previous kineme.

$$P(w_x = j|w_x^{'} = i) = \frac{\frac{1}{K} + \sum_{t=2}^{T} I(w_x(t-1), i) I(w_x(t), j)}{1 + \sum_{t=2}^{T} I(w_x(t-1), i)} \tag{7}$$
$$i = 1, 2, \cdots, K; \; j = 1, 2, \cdots, K.$$

For each clustering scheme with K clusters, we obtain $8K$ divergence values as listed in Table I (right side). One example of the conditional divergence is shown in (8). In total there are $\sum_{K=4}^{25} 8K = 2552$ divergence values.

$$D_c\big((w_x|w_x^{'} = i)\big|\big|(h_x|h_x^{'} = i)\big) = \texttt{Div}\big(P(w_x|w_x^{'} = i), P(h_x|h_x^{'} = i)\big) \tag{8}$$

V. Experiment Design and Results

A. Data split and problem setup

For each of the four behavior codes (*Acceptance*, *Blame*, *Positive*, *Negative*), we split the 221 sessions into five groups according to the expert-annotated code values of the couple. The groups of interest are: both spouses exhibit the behavior strongly (above 60*th* percentile, region I), both exhibit it weakly (below 40*th* percentile, region III), or at least one of the spouses exhibits about average (between 40-60*th* percentile, region V) on the behavior scale. The final two groups are opposite behavior traits among the couple: one exhibiting strongly and the other weakly (region II & IV). Such split is motivated by the correlation of 0.48 to 0.68 for the same

TABLE II: Count of sessions in each region and the correlation $\rho_{\texttt{code}}$ between B_w and B_h on all sessions

Code	I	II	III	IV	V	$\rho_{\texttt{code}}$
Acceptance	58	8	56	8	91	0.68
Blame	49	9	51	12	100	0.48
Positive	62	7	56	10	86	0.63
Negative	58	12	62	9	80	0.64

TABLE III: Statistics of $\mathcal{R}_d$, $\mathcal{R}_j$, $\mathcal{R}_c$, *i.e.,* the correlations between $B_w + B_h$ and $\mathbf{D}_d$, $\mathbf{D}_j$, $\mathbf{D}_c$. Note that the total number of $\mathcal{R}_d$, $\mathcal{R}_j$, $\mathcal{R}_c$ entries for each code are 44, 220, 2552, respectively. Columns are color-coded for readability.

Correlation Statistics												
	Maximum			Median			Minimum			Ratio of Sig. div.		
Code	$\mathcal{R}_d$	$\mathcal{R}_j$	$\mathcal{R}_c$	$\mathcal{R}_d$	$\mathcal{R}_j$	$\mathcal{R}_c$	$\mathcal{R}_d$	$\mathcal{R}_j$	$\mathcal{R}_c$	$\mathcal{R}_d$	$\mathcal{R}_j$	$\mathcal{R}_c$
$\texttt{Acc}_w + \texttt{Acc}_h$	0.21	0.24	0.49	0.16	0.17	0.06	0.06	0.05	−0.27	0.25	0.35	0.22
$\texttt{Bla}_w + \texttt{Bla}_h$	0.16	0.18	0.49	0.11	0.12	0.05	0.07	0.07	−0.42	0.00	0.00	0.22
$\texttt{Pos}_w + \texttt{Pos}_h$	0.37	0.41	0.54	0.33	0.35	0.16	0.26	0.26	−0.17	1.00	1.00	0.45
$\texttt{Neg}_w + \texttt{Neg}_h$	0.20	0.22	0.48	0.16	0.18	0.07	0.13	0.13	−0.38	0.11	0.50	0.31

code between the husband and wife; *i.e.,* the couple tend to behave in a similar fashion. For example, we illustrate in Fig. 3 the code values (Acc_w, Acc_h) and the corresponding split for the *Acceptance* behavior.

We train the head motion clustering model on region V (the middle cross), where at least one of the spouses has a middle range score. We carry out the analysis on region I and III, where the couple demonstrate a consistent polarity of behavioral characteristics. We do not use data from region II and IV in this study.

Note that by such split, the data used for motion model training is disjoint to that for behavior analysis. Let B_w and B_h denote the behavior code values for the wife and the husband, respectively. We take $B_w + B_h$ as the summed code for a session in region I and III, and consider the symmetric divergence in Sec. IV for the couple in each session. We denote the aggregated divergences from all the kineme clusterings and conjugate pairs obtained in Sec. IV-A, Sec. IV-B, and Sec. IV-C as $\mathbf{D}_d$, $\mathbf{D}_j$, and $\mathbf{D}_c$, respectively. The count of sessions in each region and the correlation between B_w and B_h are summarized in Table II.

We examine the following two questions:

A. Are $\mathbf{D}_d$, $\mathbf{D}_j$, $\mathbf{D}_c$ correlated with $B_w + B_h$? We denote the corresponding correlations as $\mathcal{R}_d$, $\mathcal{R}_j$, $\mathcal{R}_c$ respectively, and as $\mathcal{R}$ jointly.

B. Can $\mathbf{D}_d$, $\mathbf{D}_j$, $\mathbf{D}_c$ predict whether a session belongs to region I or region III? We denote the accuracy generally as $Acc_{\text{I,III}}$.

For **A**, we use student's t-test to test the significance of correlation. For **B**, we conduct a leave-one-couple-out cross validation (the same couple may have multiple sessions), using linear support vector machine as a binary classifier.

B. Experiment results

We hypothesize that larger divergence associates with negatively valenced emotion in communication, hence the correlation $\mathcal{R}$ would have positive values for negative emotional codes such as *Blame* and *Negative*, and have negative values for positively valenced emotional codes such as *Acceptance* and *Positive*. For ease of notation, we flip the sign of $\mathcal{R}$ for *Acceptance* and *Positive* codes, and expect $\mathcal{R}$ to be consistently larger than zero. In Table III we show the statistics of $\mathcal{R}$ after the sign flipping. Note that in the last column we compare the ratios of significant entries for $\mathcal{R}_d$, $\mathcal{R}_j$, and $\mathcal{R}_c$ at $p < 0.05$ level, while the maximum correlation achieved is much more statistically significant ($p \approx 10^{-7}$).

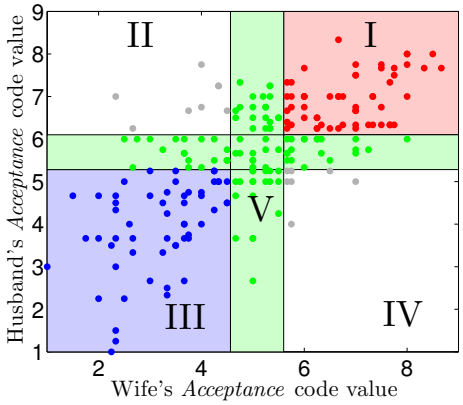

Fig. 3: Illustration of dividing the dataset into five regions. We use region V (the cross) for training head motion model, and region I, III for modeling the target behaviors.

We see that $\mathcal{R}_c$ is much higher than $\mathcal{R}_d$, $\mathcal{R}_j$ in terms of the maximum of correlations, but much lower for the median of correlations. $\mathcal{R}_c$ also has many more statistically significant results, however, in terms of the ratio it does not exceed $\mathcal{R}_d$ or $\mathcal{R}_j$ due to the difference of their sizes (except *Blame*). We will discuss the observation on the minimum values in Sec. VI. In sum, $\mathcal{R}_c$ contains strong but sparse (in terms of ratio) entries of significant correlations. $\mathcal{R}_d$ and $\mathcal{R}_j$ are consistent and modest in significance.

Furthermore, the result varies depending on the behavior code. For $\mathcal{R}_d$ and $\mathcal{R}_j$, the correlations are relatively stronger

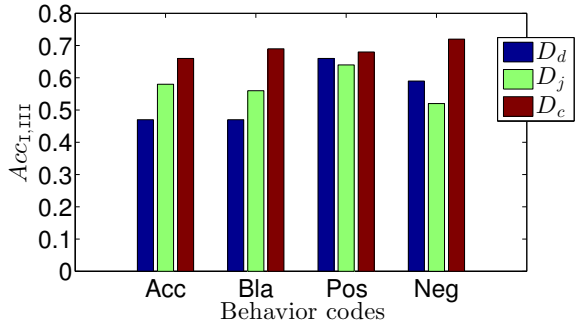

Fig. 4: Accuracies of region I *vs.* III classification.

for code of general emotion and attitude such as *Positive* and *Negative*, but less prominent for code *Blame*, which is primarily manifested through the verbal channel. However, the statistics of the correlations are comparable for all codes with respect to $\mathcal{R}_c$. This suggests that conditioning on particular kinemes may help to capture behavioral similarity in a given context, which may be more informative for inferring subject's behavioral characteristics such as *Blame*.

In Fig. 4 we show the classification performance $Acc_{\mathrm{I,III}}$ in problem **B**. D_d yields better than chance classification for some codes, but is not effective in general. D_j is comparable in general to D_d. D_c outperforms both D_d and D_j, improving $Acc_{\mathrm{I,III}}$ appreciably for some codes. This suggests D_c, the conditional divergence of behavior, is better suited than the other two in inferring behavioral characteristics.

Comparisons of the performance by D_c to that of the other modalities such as acoustic [32], lexical [33], and head motion types [20] are not strictly applicable, due to mismatch of experiment settings. Nevertheless, the results in this work may suggest that head motion similarity should be considered as one useful stream of cues informing human behavioral characteristics.

Finally, in order to analyze the effect of the number of kineme types K, we compute the mean-absolute values of $\mathcal{R}$ over all codes for different K, denoted $\overline{|\mathcal{R}|}^K$. The mean and standard deviations of $\overline{|\mathcal{R}|}^K$ with respect to K are listed in Table IV. These relatively small standard deviations suggest that the correlations represented in $\mathcal{R}$ are not very sensitive to the choice of K.

TABLE IV: Mean and standard deviations of $\overline{|R_d|}^K$, $\overline{|R_j|}^K$, $\overline{|R_c|}^K$, with respect to $K \in \mathbf{N}^+$, $4 \leq K \leq 25$

Std.	*Acceptance*	*Blame*	*Positive*	*Negative*
$\overline{\lvert R_d\rvert}^K$	0.15±0.02	0.11±0.02	0.33±0.01	0.16±0.01
$\overline{\lvert R_j\rvert}^K$	0.17±0.02	0.12±0.01	0.35±0.02	0.18±0.01
$\overline{\lvert R_c\rvert}^K$	0.11±0.01	0.13±0.01	0.18±0.02	0.14±0.01

VI. Discussion: saliency of kinemes

We explore if the conditioning to particular kinemes in D_c associates with higher value entries in $\mathcal{R}_c$, *i.e.,* if some specific kinemes are more salient stimuli in human communication. Recall that D_c is obtained by the divergence of current kineme distributions conditioned on the previous kineme, where the latter may have varying degrees of importance in behavior modeling. We first employ the entries with significant $\mathcal{R}_c$ ($p < 0.01$) in D_c (1798 in total for all codes), and look for the label i of the conditioned kineme. We then compute the correlation ρ_i between the count of kineme i per session per subject and the corresponding behavior code, *e.g.,* the correlation between the count of $w_x(t) = i$ in a session and B_w. Thus we check if ρ_i is also significant ($p < 0.01$) for at least one spouse. Such significance may indicate kineme type i as salient behavioral cue in expressing affect and attitude. As a result, we find that in average for all codes, 68% of significant entries in D_c are conditioned on a kineme that also has significant ρ_i. This suggests that the majority of salient conditional divergences are obtained via the behavior descriptors that are temporally conditioned on a salient kineme.

In Table III the minimum values of $\mathcal{R}_d$ and $\mathcal{R}_j$ are above zero. This suggests that despite non-significant entries, all the entries are in the consistent polarity of larger divergence indicating more negative relationship. However, we observe highly negative correlation values in $\mathcal{R}_c$. Similar to the above operations, we collect entries of $\mathcal{R}_c < 0$ with $p < 0.01$, look for the conditioned kineme label i, and compute ρ_i. As a result, we find 71 significant cases for all codes, out of which 70 cases are conditioned on a kineme of $\rho_i > 0$ with negative emotional codes (or $\rho_i < 0$ for positive emotional codes) for both spouses, *i.e.,* destructive to relationship. Among the above 70 cases, 49 cases have significant ($p < 0.01$) ρ_i for at least one spouse. This suggests that *dis*-entrainment to destructive stimuli may associate with positive emotion. Such results lend support to the *disconnection* function of breaking entrainment, which sends the message that "we are *not* on the same page" [1].

VII. Conclusion

In this work we proposed head motion similarity measures as approximation to behavioral entrainment in couples' dyads. We employed a data-driven clustering model that transcribes head motions into kineme streams. We proposed three groups of similarity measures of the couples' interaction computed using Kullback-Leibler Divergence. We found that the correlations between the similarity measures based on the (joint) distribution of kinemes and the behavioral characteristics were consistent and significant, although modest. The similarity measures based on the conditional distribution of kinemes had strong, albeit sparse, correlations with the couples' behavioral characteristics, and were most effective in predicting the polarity of couples' behavioral characteristics. The majority of strong correlations were obtained when the corresponding distributions for the similarity measures were conditioned on kinemes with salient behavioral implications. In some cases, larger divergence indicated positive emotion despite that the vast majority was in the opposite as commonly suggested in theory. We found that in these cases the distributions were conditioned on kinemes implying negative emotion; these results offer corroborating evidence for the disconnection function of dis-entrainment. In sum, we have demonstrated the feasibility of inferring behavioral characteristics through the modeling of head motion entrainment.

In the future we would like to train the head motion model independent of each code. Thus we can investigate the behavioral meaning of the same kineme with respect to different behavioral codes, both from an engineering viewpoint and from the insight of psychologists. We would also like to further study the salient kinemes; we plan to work both through the data-driven approach and also employ domain experts to identify these salient behaviors. This would enable algorithms to detect and model such salient instances as well as study the impact of them on the behavior of the interlocutors and in quantifying the interlocutors' entrainment. In addition, we would like to study further the multimodal nature of the signals in such interactions.

REFERENCES

[1] T. Wheatley, O. Kang, C. Parkinson, and C. Looser, "From mind perception to mental connection: Synchrony as a mechanism for social understanding," *Social and Personality Psychology Compass*, vol. 6, no. 8, pp. 589–606, 2012.

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

[3] M. Iacoboni, "Imitation, empathy, and mirror neurons," *Annual review of psychology*, vol. 60, pp. 653–670, 2009.

[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, p. 284, 2011.

[5] F. Bernieri, J. 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, p. 243, 1988.

[6] F. 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] D. Julien, M. Brault, É. Chartrand, and J. Bégin, "Immediacy behaviours and synchrony in satisfied and dissatisfied couples." *Canadian Journal of Behavioural Science*, vol. 32, no. 2, p. 84, 2000.

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

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

[11] E. Delaherche, M. Chetouani, A. Mahdhaoui, C. Saint-Georges, S. Viaux, and D. Cohen, "Interpersonal synchrony: A survey of evaluation methods across disciplines," *IEEE Transactions on Affective Computing*, vol. 3, no. 3, pp. 349–365, 2012.

[12] 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.

[13] D. Sander, D. Grandjean, S. Kaiser, T. Wehrle, and K. R. Scherer, "Interaction effects of perceived gaze direction and dynamic facial expression: Evidence for appraisal theories of emotion," *European Journal of Cognitive Psychology*, vol. 19, no. 3, pp. 470–480, 2007.

[14] M. Valstar, J. Girard, T. Almaev, G. McKeown, M. Mehu, L. Yin, M. Pantic, and J. Cohn, "Fera 2015-second facial expression recognition and analysis challenge," *Proc. FG*, 2015.

[15] Z. Hammal, J. F. Cohn, D. S. Messinger, W. Mattson, M. H. Mahoor *et al.*, "Head movement dynamics during normal and perturbed parent-infant interaction," in *Proc. ACII*. IEEE, 2013, pp. 276–282.

[16] Z. Hammal, J. F. Cohn, and D. T. George, "Interpersonal coordination of head motion in distressed couples," *IEEE Transactions on Affective Computing*, vol. 5, no. 2, pp. 155–167, 2014.

[17] E. McClave, "Linguistic functions of head movements in the context of speech," *Journal of Pragmatics*, vol. 32, no. 7, pp. 855–878, 2000.

[18] R. Birdwhistell, *Kinesics and context: essays on body motion communication*. University of Pennsylvania Press, 1970, vol. 2.

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

[20] B. Xiao, P. G. Georgiou, B. Baucom, and S. S. Narayanan, "Power-spectral analysis of head motion signal for behavioral modeling in human interaction," in *Proc. ICASSP*. IEEE, 2014, pp. 4593–4597.

[21] B. Xiao, P. Georgiou, C. Lee, B. Baucom, and S. Narayanan, "Head motion synchrony and its correlation to affectivity in dyadic interactions," in *Proc. ICME*, 2013.

[22] A. Christensen, D. Atkins, S. Berns, J. Wheeler, D. Baucom, and L. 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.

[23] C. Heavey, D. Gill, and A. Christensen, "Couples interaction rating system 2 (CIRS2)," *University of California, Los Angeles*, 2002.

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

[25] F. Ramseyer and W. Tschacher, "Synchrony: A core concept for a constructivist approach to psychotherapy," *Constructivism in the human sciences*, vol. 11, no. 1, pp. 150–171, 2006.

[26] R. G. Reed, A. K. Randall, J. H. Post, and E. A. Butler, "Partner influence and in-phase versus anti-phase physiological linkage in romantic couples," *International Journal of Psychophysiology*, vol. 88, no. 3, pp. 309–316, 2013.

[27] J. Butner, L. M. Diamond, and A. M. Hicks, "Attachment style and two forms of affect coregulation between romantic partners," *Personal Relationships*, vol. 14, no. 3, pp. 431–455, 2007.

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

[29] J. Shi and C. Tomasi, "Good features to track," in *Proc. CVPR*. IEEE, 1994, pp. 593–600.

[30] J.-Y. Bouguet, "Pyramidal implementation of the affine lucas kanade feature tracker description of the algorithm," *Intel Corporation*, vol. 5, 2001.

[31] U. Hadar, T. Steiner, E. Grant, and F. Rose, "Kinematics of head movements accompanying speech during conversation," *Human Movement Science*, vol. 2, no. 1, pp. 35–46, 1983.

[32] M. Black, A. Katsamanis, B. B.R., C. Lee, A. Lammert, A. Christensen, P. Georgiou, and S. Narayanan, "Toward automating a human behavioral coding system for married couples interactions using speech acoustic features," *Speech Communication*, vol. 55, no. 1, pp. 1–21, 2013.

[33] P. Georgiou, M. Black, A. Lammert, B. Baucom, and S. Narayanan, ""that's aggravating, very aggravating": Is it possible to classify behaviors in couple interactions using automatically derived lexical features?" in *Proc. ACII*, 2011, pp. 87–96.