Is Interactional Dissynchrony a Clue to Deception: Insights from Automated Analysis of Nonverbal Visual Cues

Xiang Yu[1], Shaoting Zhang[1], Zhennan Yan[1], Fei Yang[1], Junzhou Huang[1],
Norah Dunbar[3], Matthew Jensen[3], Judee Burgoon[2], Dimitris Metaxas[1]

Computer Science Department, Rutgers University, Piscataway, NJ 08854
Center for the Management of Information, University of Arizona, Tucson, AZ 85721
Center for Applied Social Research, University of Oklahoma, OK 73019
{xiangyu, shaoting, zhennany,feiyang,dnm}@cs.rutgers.edu, {ndunbar,mjensen}@ou.edu, jburgoon@cmi.arizona.edu

Abstract

In this paper, we investigate how degree of, and temporal changes in, interactional synchrony can signal whether an interactant is truthful or deceptive. We propose an automated, data-driven and unobtrusive framework using visual cues. Our framework consists of face tracking, gesture detection, facial expression recognition and interactional synchrony estimation. This framework is able to automatically track gestures and expressions of both the target interviewee and the interviewer, extract normalized meaningful synchrony features and learn classification models for deception recognition. To validate these proposed synchrony features, extensive analyses have been conducted on a database of 242 video samples and show that these features reliably capture simultaneous synchrony. The relationship between synchrony and deception is shown to be complex.

1. Introduction

Implicit in all interpersonal interactions is the need to gauge whether an interlocutor is truthful and authentic in his or her presentation of self. The expectation of truthfulness, in fact, is one of the foundations of human discourse [11]. Yet, notwithstanding the importance of this largely outside-of-consciousness assessment process, voluminous research has shown that humans, unaided by technology, are very poor at detecting deception [1, 2, 19]. Average detection accuracy is estimated at 54%, or slightly above chance, and detection of deception specifically, as opposed to detection of truthfulness, is approximately 47% [2]. Those accuracy estimates have included both lay and professional judges, although some recent evidence points to experts achieving higher accuracy rates under interviewing conditions more characteristic of their usual professional setting and task [5].

One reason cited for humans' poor detection in interpersonal dialogue is that deceivers take advantage of the give-and-take of interaction to adapt to any signs of skepticism in the interviewer's verbal and nonverbal feedback. Deceivers adjust their messages to make their responses more plausible and their demeanor more credible [3, 20]. That same give-and-take, however, has the potential to offer subtle clues to deception through the disruption of interactional synchrony. Interactional synchrony refers to interaction that is non-random, patterned, and aligned in both timing and form. Kinesic behavior (i.e., head, face, body and limb movement) is coordinated to the rhythms and forms of expression in the vocal-verbal stream. It is considered a key marker of interaction involvement, rapport, and mutuality. It may take the form of simultaneous synchrony, in which two or more people's behaviors mimic or match one another (e.g., similar postures and facial expressions) in the same time frame and behavioral changes occur at the same junctures. This is speaker-listener synchrony. Synchrony may also be concatenous, in which one person's behavior while speaking is followed by similar behavior from the next speaker (e.g., each using rapid nodding while speaking). This serial form of synchrony captures speaker-speaker and listener-listener coordination.

Hypothesis: The current investigation explores simultaneous synchrony. It is premised on the possibility that engaging in deception disrupts interactional synchrony and may therefore be a clue to its presence. Practitioners have suggested using rapport-building techniques or interactional synchrony as an effective method for detecting deception: with terrorists in FBI interviews, and in police investigations [13, 16, 17]. However, few systematic studies of rapport, coordination, synchrony or reciprocity have examined the effects of synchrony on deception or vice versa [4, 6]. The emphasis typically has been on interviewers using interactional synchrony to promote more verbal disclosures and confessions by interviewees.

Our approach is a novel perspective on the role of synchrony in revealing deception in that we are focusing on the interaction between the interviewee and interviewer rather than only the interviewer side of the equation. Deception has been shown to be a cognitively and emotionally taxing activity, especially when the stakes are high and the consequences of being discovered are serious [12, 18]. Interactional synchrony entails a very close linkage among behavioral, physiological and emotional synchrony such that synchrony is positively correlated with rapport and empathy between interlocutors; conversely, incongruent feeling states and behavioral states can disrupt coordination, synchrony and perceived rapport [15]. Because deceivers may experience various negative emotional states (or at least emotional states that diverge from those of an interlocutor) and because deceivers may be too preoccupied with constructing plausible verbal responses to attend to or coordinate their nonverbal behaviors with another, we expect interactional synchrony to be attenuated and disrupted when interviewees are deceptive as compared to when they are truthful. Even skilled deceivers may be unable to counter this decrement in interactional enmeshment because conscious efforts to produce synchronous behavior patterns through mirroring another's posture or matching their degree of animated gesturing and facial expressions may appear "inauthentic" and "off" [10]. Our hypothesis tested this possibility. Deception thus may be one cause of poor interactional dissynchrony and dissynchrony may be one sign that deception is taking place.

Our assumption is interviews with deceivers are less synchronous than interviews with truth tellers. Testing this hypothesis required developing the computer vision methods to assess simultaneous synchrony. These methods are the central focus of the current report.

Moderators: Little is known about whether moderator variables alter the patterns of synchrony. Three possible factors were investigated here: the modality of interaction (face-to-face or video-conferencing), sanctioning of the deception, and nature of the topic or interview questions. These variables and the reason for their selection for this experiment are detailed elsewhere [7]. Few experiments have examined videoconferencing and instead compare face-to-face interactions to those in text-only modalities [9]. In addition, few experiments directly compare the situation where the experimenter has sanctioned the deception to unsanctioned deception [8] and instead tend to focus on one or the other. In many experiments, participants are told by an experimenter to deceive their partner which may result in less nervousness, guilt, and dissynchrony. In other experiments, participants are allowed to choose whether or not to lie, which results in a lack of random assignment with only confident or skilled deceivers choosing unsanctioned deception.

A third likely moderator is the topic or question under discussion. It stands to reason that differing degrees of involvement in a topic, its sensitivity, and its potential for placing an interlocutor in an embarrassing or compromised position might sever the degree of coordination between two people. For example, talking about mundane topics while two people attempt to establish some common ground and to jointly create a comfortable interaction may lead to the development of rapport and synchrony whereas asking a question that implies some challenge to another's honesty and veracity is likely to destabilize interaction rhythms as the person under question seeks to make sense of the direction the conversation is heading and reduce his or her own disquiet and uncertainty. Interactional synchrony is fluid and, as a joint product of each party's emotional, cognitive and physiological states, is bound to change across phases of an interaction. Characterizations of an entire interaction as synchronous or dissynchronous gloss over the possibility of temporal variability in the degree of synchrony as a function of what is being discussed.

Our experiment examines all three of these moderators, modality, sanctioning and topic to determine the extent to which they disrupt or encourage synchrony between interviewer and interviewee. We asked the following research questions: (1): Is the synchrony between interviewer and interviewee affected by the modality they are using (face-to-face or videoconferencing)? (2): Does the sanctioning of the deception (sanctioned or unsanctioned by experimenter) affect the synchrony between interviewer and interviewee?

Method: In overview, our approach was as follows. Stimulus videos were derived from a cheating experiment in which some subjects cheated during a trivial pursuit game and some did not, but all were encouraged to appear as credible as possible when interviewed about the game. Thus cheaters were expected to be deceptive and non-cheaters, to be truthful. Modality and sanctioning were experimentally manipulated such that some participants were interviewed face-to-face and others were interviewed with computer-mediated communication via Skype. Some were told that the experimenters were aware of their cheating but that they were to deny it to the interviewer (sanctioned version) whereas others received no such explicit approval or their cheating (unsanctioned cheating).

Interviews were conducted by certified professional examiners supplied by a federal agency and included three phases of questioning: an initial baseline set of questions that were benign, a set of questions that presented indirect accusations, and a final set that directly inquired about cheating. Analyses were conducted according to these three phases of the interview.

The videotaped interviews were analyzed using two computer vision methods for automated analysis, supplemented by manual coding. Skin Blob Tracking was used to track gross body movements (posture, gestures, head movements) and Active Shape Modeling was used for detailed face tracking. Additionally, trained human coders conducted behavioral observation to (1) rate immediacy behaviors, (2) record changes in immediacy and (3) test synchrony between two partners' change in immediacy levels. The results of the manual human coding are reported elsewhere (Dunbar, et al., 2011). Our focus in this report is on the automated tracking of gestures and expressions of both the subject and the interviewer, extracting normalized meaningful synchrony features and learning classification models for deception recognition.

2. Methodology

We have developed a framework that is capable of tracking facial movements and detecting the level of synchrony in real time, as shown in Figure 1. Our framework consists of four components: face tracking, gesture detection, expression recognition, and synchrony estimation. We will introduce each component in this section.

2.1. Multi-pose Face Tracking

Face tracking is a challenging problem. The shapes of faces change dramatically with various identities, poses and expressions. Furthermore, poor lighting conditions may cause a low contrast image or cast shadows on faces, which will significantly degrade the performance of the tracking system. We have developed a robust face tracker [21] based on Active Shape Models (ASMs) [22] together with a nonlinear shape manifold.

Figure 1: Sample snapshots from tracked facial data showing a subject (left) and an interviewer (right). Red dots represent tracked facial landmarks (eyes, eyebrows, etc.), while ellipse in top left corner depicts the estimated 3D head pose of the subject; top right corners show the detected expressions and head gestures for subject and interviewer.

ASMs are landmark-based models that attempt to learn a statistical distribution over variations in shapes for a given class of objects. The ASM consists of a global shape model and a set of local landmark detectors. The global shape model captures shape variations, whereas the local profile models capture the local appearances around each landmark point and are used for selecting the best candidate landmark positions. To locate the facial features in varying poses, we learn a group of shape models, each covering a range of face poses. At each frame the system traverses the non-linear facial shape manifold looking for the landmark configuration whose shape and texture at each landmark yield the minimum distance between what is observed in the image and the reconstructed shape. As show in Figure 2, the learned model allows the complex, non-linear shape manifold to be approximated in piecewise linear sub-regions. Each sub-region defines a hyper-ellipsoid on this manifold. Facial shapes of similar pose are constrained to lie in the same linear subspaces. We have also employed sparse shape representation [23, 24] to model varying poses. This method models facial shapes as a sparse linear combination of training shapes. Therefore, it is able to model multi-distribution of training shapes, i.e., varying poses in this application.

2.2. Gesture and Facial Expression Detection

Using the landmark positions obtained from our face tracking method, we are able to estimate the 3D poses (pitch, yaw, and tilt) and detect the relevant gestures (head shaking and nodding). To estimate the face pose, we built a linear regression model for each linear region in the shape manifold. The regression model takes the X and Y coordinates of the 79 landmarks as input, and predicts the pitch, yaw and tilt angles.

The face nodding is rapidly and repeatedly moving the face up and down. By differentiating the pitch value in each frame, we are able to detect the head nodding and shaking. Also we have built a facial expression classifier to detect smiles. Our method uses the relative intensity ordering of facial expressions found in the training set to learn a ranking model (RankBoost) [25]. We extract the haar-like features to represent facial appearance, and use the RankBoost to select a subset of haar-like features to build a final classifier. Our method not only recognizes a specific facial expression, but also estimates the intensity of facial expression.

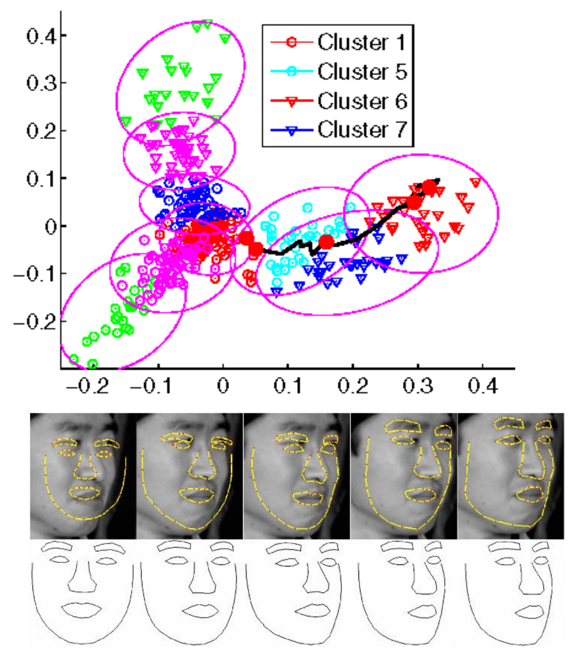

Figure 2: **Top**: The face shape manifold is approximated by piecewise linear sub-regions.
Bottom: Our method searches across multiple clusters to find the best local linear model.

2.3. Synchrony Features

The subtle and significant way people influence each other can be seen through their nonverbal synchrony. Synchrony refers to similarity in rhythmic qualities and enmeshing or coordination of the behavioral patterns of both parties in an interaction [26]. Such synchrony can either be simultaneous or concatenous. In Dunbar et al.'s work [6], synchrony is reflected from gesture, nodding or shaking, facial mirroring, etc. When providing pairs of interview videos, we could obtain head nodding or shaking and facial expression especially smiling information of people in the videos by our proposed facial tracking and facial expression detection methods. Based on such lower level features, we intend to check the simultaneous or concatenous response from both people in one interview.

Lower level feature vectors of two interview videos from one interviewer and one interviewee can be viewed as two corresponding data sequences. We know that we can obtain large responses while correlating two sequences if the two sequences have similar magnitude at the same position. This can measure the simultaneous response. If two sequences have similar magnitude at different positions, we can take a time sliding window for the time delay and then calculate their correlation. Cross correlation is a standard method of estimating the degree to which two sequences are correlated. The definition of cross correlation of two signal series of which one is delayed at gap d is as:

$$C(d) = \frac{\sum_i (x_i - \bar{x})(y_{i-d} - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2}\sqrt{\sum_i (y_{i-d} - \bar{y})^2}} \tag{1}$$

where x_i and y_i are the i^{th} element of sequence x and sequence y, $\bar{x}$ and $\bar{y}$ are the mean value of x and sequence y.

In order to estimate concatenous synchrony, we divide two sequences into overlapped time slots which can be seen in Figure 3. Firstly the two sequences are required to have the same length. Then we equally divide each sequence into m time slots. Starting from either of the sequences, for the current time slot, we go backward t time slots and forward of t time slots to calculate their correlation with the current time slot. We choose the largest cross correlation value as the current time slot's feature value. We repeat such procedure for every time slot in one sequence until the end of the sequence. Thus we obtain a cross-correlation based higher level feature vector with length m.

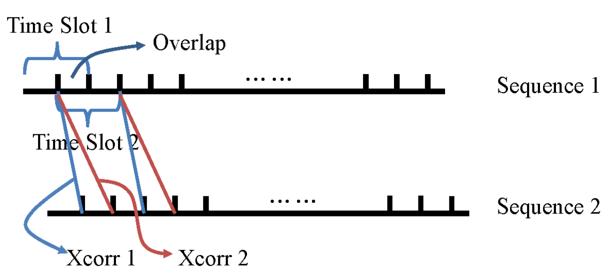

Figure 3: sequences cross correlation scheme

3. Experiments

The analysis began with creating a database of 242 videotaped interviews of 121 interviewer-interviewee pairs. Interviewees were students who participated in a trivial game and in some cases were induced by a confederate to cheat. All participants were then interviewed by expert interviewers about the game interaction and whether they cheated during the game. Approximately half of the interviews were conducted over Skype and the other half were conducted face-to-face to produce two modality conditions, Computer-Mediated Communication (CMC) and Face to Face Communication (FtF). Since a few of the pairs are incompletely recorded, we finally pick 100 out of 121 pairs of videos as our training and testing data. These video pairs vary from 4500 frames to 15000 frames. Although video pairs' lengths are different, we ensure inside each video pair, the interviewer sequence and interviewee sequence keep the same length, which allows using fixed number of time slots to analyze the video sequences.

In the synchrony detection step, we have extracted head nodding, head shaking, smiling and head positions facial gestures. Based on such lower level features, we further combine the interviewer's feature vector with interviewee's feature vector to form higher level features. Correlation based method is adopted to identify synchrony. Then a tree structure classification scheme is designed to separate 3 classes (truthful group, sanctioned cheating group and unsanctioned cheating group). We first aim to classify the truthful group from the cheating group using non-linear SVM classifier, which is a two class classification task. Then based on the result of the first step, we continue to classify the cheating group into sanctioned cheating group and unsanctioned cheating group by another non-linear SVM classifier. During the feature selection part, at each step we separately train a feature selector using Genetic Algorithms. The feature selector is an efficient way to promote the performance in recognition task because the raw features may have noise or redundancy.

3.1. Evaluation of Features

Before sending features into classifiers, the different types of features should be investigated as to which ones are effective for classification. Our major strategy is to leave each single feature out of the whole feature vector and then test the recognition accuracy. Also, we identifies single feature's recognition accuracy and visualize the feature vector in plots to see the distinguishability of each of the four types of features (head nodding, head shaking, smiling or not smiling and look forward or look away). During this step, we examine the average precision of classifying the three classes of the sanctioning variable (truthful, sanctioned cheating and unsanctioned cheating classes) to evaluate each feature vector. Table 1 shows the average precision of different feature combinations over the three-class classification.

Table 1: Evaluation of four features, i.e., “Nod”, “Shake”, “Smile” and “Look forward”. “All but one” means that all features are used except the one of that column. “Single” means using only the feature of that column.

		Nod	Shake	Smile	Look forward
CMC	All but one	0.422	0.437	0.356	0.311
	Single	0.35	0.33	0.38	0.43
FtF	All but one	0.442	0.473	0.554	-
	Single	0.513	0.464	0.435	-

From Table 1, we see that for CMC, when the feature "Nod" or "Shake" is excluded from the whole feature vector, the performance is higher than the rest. When the feature "Smile" or "Look forward" is excluded, the performance drops. For FtF, the trend is opposite: “Nod” and “Shake” are more significant in classification. When testing each single feature's accuracy, it appears that "Look forward" and "Smile" are more accurate than "Nod" and "Shake" for CMC. And again, for FtF modality, “Nod” and “Shake” achieves higher accuracy than “Smile”, where “Look forward” is not applicable in FtF dataset. The reason is for FtF data, interviewer and interviewee sit face to face. The look-forward feature should be defined by their local head coordinates. But only one camera was capturing the single scene, which allows just the global camera coordinate. We cannot get the frontal face by the camera coordinate.

In Figure 4, the vertical dot lines separate the plot into 4 regions representing the four separate features. The first column indicates feature "Nod", the second one is feature "Shake", the third is feature "Smile" and the last one is feature "Look forward". We plot the average feature vector of each group in the subplots. The feature vector is 800 numbers long, of which each region is with length 200 numbers. With black line showing the trends in the figure, we see that in region three, the pattern of the feature vector is obviously different. In the subplot for the truthful condition, it is going down; in the subplot for unsanctioned cheating, it is flat; in subplot of sanctioned cheating, it is going up. In region four, the average value of those numbers is going down from above 0.9 to less than 0.9 until around 0.8.

3.2. Evaluation of the Two-Class Classification

The initial 3 class classification using non-linear SVM scheme may not be perfect because it contains at least 3 intersections of misclassification, which are intersections of truthful and unsanctioned cheating groups, truthful and sanctioned cheating groups, unsanctioned cheating and sanctioned cheating groups. Although the problem is to divide the data into truthful, unsanctioned cheating and sanctioned cheating groups, it is at least a 2 classes' classification problem, of which is truthful and cheating groups' classification. We could continue to solve a 2 classes' classification problem on the unsanctioned cheating and sanctioned cheating groups in the same way. Hence, we get only 2 intersections of misclassification, misclassification of truthful and cheating groups and misclassification of unsanctioned cheating and sanctioned cheating groups, which is expected to decrease the error recognition rate. We set both 15 test samples for truthful group and cheated group. Thus 70 samples are the training samples, 16 in the truthful group and 54 in the cheating group. The performance is shown in Table 2.

Table 2: The confusion matrices of classifying truthful and deceptive cases of CMC and FtF modalities.

		Truthful	Deceptive
CMC	Truthful	10	5
	Deceptive	6	24
FtF	Truthful	12	4
	Deceptive	7	20

Table 3: The accuracy of classifying the truthful and deceptive cases. “TP” and “FP” stand for true positive and false positive.

		TP	FP	Precision	Recall
CMC	Truthful	0.667	0.200	0.625	0.667
	Deceptive	0.800	0.333	0.828	0.800
	Average	0.734	0.267	0.727	0.734
FtF	Truthful	0.75	0.259	0.632	0.75
	Deceptive	0.741	0.25	0.833	0.741
	Average	0.744	0.253	0.758	0.744

The confusion matrices in Table 2show that for the CMC dataset in the truthful group, 10 samples are correctly classified while 5 are not; in the cheating group, which is the combination of unsanctioned cheating and sanctioned cheating groups, 24 samples are correctly classified and only 6 are not. Table 3 shows the classification accuracy details. In CMC, “truth” precision is 0.625 and “deception” precision is .082, for an overall average of .727. In the FtF condition, the precision values are .632 for “truth” and .833 for “deception” for an overall precision of 0.758, which is at the same level of the CMC dataset.

3.3. Evaluation of the Three-Class Classification

The next step in the analysis was to classify the cheating group into unsanctioned cheating and sanctioned cheating groups. In this case, the training and test sets are data-dependent, especially in feature selection and non-linear SVM classifier training.

Table 4 shows our final confusion matrices over all the three categories. In each category, the number of correctly recognized samples dominates misclassified numbers.

Table 4: The confusion matrices of classifying truthful, unsanctioned and sanctioned cheating cases of CMC and FtF modalities, “Unsanctioned” stands for unsanctioned cheating and “Sanctioned” stands for sanctioned cheating.

		Truthful	Unsanctioned	Sanctioned
CMC	Truthful	10	2	3
	Unsanctioned	4	9	2
	Sanctioned	2	2	11
FtF	Truthful	11	1	4
	Unsanctioned	5	7	1
	Sanctioned	2	2	10

Further Table 5 illustrates that almost all the three classes' precisions are above 0.6, two of which are approaching 0.7. The average accuracy is 0.668, which is clearly a significant improvement over 0.47 [2].

Table 5: The accuracy of classifying the truthful, unsanctioned and sanctioned cheating cases. “TP” and “FP” stand for true positive and false positive. “Unsanctioned” and “Sanctioned” stand for unsanctioned cheating and sanctioned cheating groups.

		TP	FP	Precision	Recall
CMC	Truthful	0.667	0.200	0.625	0.667
	Unsanctioned	0.600	0.133	0.692	0.600
	Sanctioned	0.733	0.167	0.688	0.733
	Average	0.667	0.167	0.668	0.667
FtF	Truthful	0.688	0.296	0.579	0.688
	Unsanctioned	0.538	0.067	0.778	0.538
	Sanctioned	0.714	0.172	0.667	0.714
	Average	0.651	0.187	0.668	0.651

3.4. Evaluation of confessors in deception detection

In the experiment, some of the interviewees may confess to their deceiving during the interview. Before they confess to the interviewer, the pattern may appear the same as cheating mode. After the confession, they may feel relieved and then perform as truthful mode. We are suspecting that such confessor group inside cheating group may influence the detection. We want to see whether the confessors were more synchronous than the non-confessors by evaluating the excluded dataset performance. Comparing to section 3.3’s result, we expect to find the degree of synchrony from including and excluding such confession group.

Table 6: The confusion matrices of classifying truthful, unsanctioned and sanctioned cheating cases of CMC and FtF modalities in confession group excluded condition, “Unsanctioned” stands for unsanctioned cheating and “Sanctioned” stands for sanctioned cheating.

		Truthful	Unsanctioned	Sanctioned
CMC	Truthful	27	3	1
	Unsanctioned	10	18	2
	Sanctioned	4	6	7
FtF	Truthful	14	2	0
	Unsanctioned	2	2	1
	Sanctioned	2	1	3

Table 6 reports the confusion matrices of the three-class classification result on both CMC and FtF databases. The diagonal elements of the two matrices dominate all the other elements which reflect that our classification scheme groups most of the samples correct. Further comparing Table 7 with Table 5, we see that the excluding confessor classification achieves at least as good result as the including confessor scheme. Moreover, we notice that not only for CMC but also for FtF, in “Truthful” group, the excluding scheme achieves 0.806 accuracy for CMC and 0.875 accuracy for FtF, while the including scheme achieves 0.667 for CMC and 0.688 for FtF. We found a significant improvement of classifying truthful group from the deceptive group by more than 15%. The important strategy is to remove the confessor group from the deception class, which identifies our assumption that confessor group influences the synchrony degree in deception detection.

Table 7: The accuracy of classifying the truthful, unsanctioned and sanctioned cheating cases in confession group excluded condition. “TP” and “FP” stand for true positive and false positive. “Unsanctioned” and “Sanctioned” stand for unsanctioned cheating and sanctioned cheating groups.

		TP	FP	Precision	Recall
CMC	Truthful	0.871	0.298	0.659	0.871
	Unsanctioned	0.600	0.188	0.667	0.600
	Sanctioned	0.412	0.049	0.700	0.412
	Average	0.667	0.201	0.671	0.667
FtF	Truthful	0.875	0.364	0.778	0.875
	Unsanctioned	0.4	0.136	0.4	0.4
	Sanctioned	0.5	0.048	0.75	0.5
	Average	0.704	0.251	0.702	0.704

3.5. Discussion

The analyses of the CMC and FtF conditions were analyzed in parallel fashion, but the four features had different significance in the two conditions, possibly due to the physical location of the camera or individuals in each. Both in two-class and three-class classification, the performance of CMC and FtF datasets achieved the same degree of accuracy, which suggests that the degree of synchrony was not influenced too much by the modality of communication. Nevertheless, from the three-class classification result, we could see that the sanctioned cheating group is well separated from the unsanctioned cheating group, which indicates sanctioning is a key factor to influence synchrony degree and as a result discriminates unsanctioned cheating from sanctioned cheating. Those deceivers with confession during the interview influence the classification process. Once the confessor group is removed, the truth tellers are much better separated than before.

Automatic methods can often detect events of synchrony which are missed by the human coders for whatever reason. In particular, we found that the human coders would label a given video as having no synchrony in it, while our software did detect a number of synchrony events, producing disagreement between the results of the manual analysis and the results of the automatic analysis. Despite a small percentage of false negatives in detecting the events of interest (i.e., nodding, shaking, smiling), the results of the automatic analysis are supportive of the initial hypothesis of synchrony being detectable and discriminating among conditions. This means that monitoring synchrony events, while establishing implicit models of deception, may be useful for automatic deception detection.

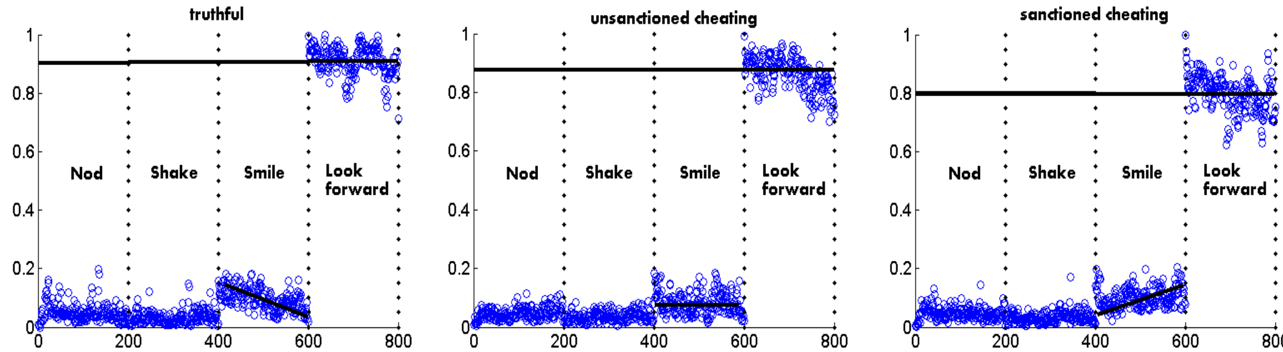

Figure 4: Mean feature vector patterns of three groups. **Left**: truthful group's mean feature vector pattern. **Middle**: unsanctioned cheating group's mean feature vector. **Right**: sanctioned cheating group's mean feature vector.

False negatives (for shaking and nodding) are attributed to the poor resolution of the input video and to the fact that the camera was not frontal to the faces. In particular, the face was quite small, and although it was correctly tracked, the displacement of the facial landmarks was sometimes not large enough to register as a nodding or shaking event. We believe that using videos of better quality, with facial close-ups, will improve our results and confirm our findings.

4. Conclusions

In this paper, we investigate how degree of interactional synchrony can signal whether deceit is present, absent, increasing or declining. An automated framework has been introduced to analyze videos effectively, and a new group of features has been proposed that not only register synchrony but also can detect deception at a reasonable accuracy. Future analysis will consider if the trend discovered thus far by our computerized methods generalizes to the greater sample population. Furthermore, we will improve our system by incorporating 3D deformable models [27, 28, 29] and sparsity based shape priors [23, 24].

References

[1] Aamodt, M. G. and Custer, H. Who can best catch a liar?: A meta-analysis of individual differences in detecting deception. *The Forensic Examiner, 15*(1), 6-11, 2006.

[2] Bond, C. F., Jr. and DePaulo, B. M. Accuracy of Deception Judgments. *Personality and Social Psychology Review, 10*(3), 214-234, 2006.

[3] Buller, D. B. and Burgoon, J. K. Interpersonal deception theory. *Communication Theory, 6*, 203-242, 1996.

[4] Burgoon, J. K., Stern, L. A. and Dillman, L. *Interpersonal adaptation: Dyadic interaction patterns*. New York: Cambridge University Press, 1995.

[5] Burgoon, J. K., Nunamaker, J. F., Jr. and Metaxas, D. *Noninvasive measurement of multimodal indicators of deception and credibility.* Final Report to the Defense Academy for Credibility Assessment, July 10, 2010.

[6] Dunbar, N. E., Jensen, M. L., Burgoon, J. K., Adame, B., Robertson, K. J., Harvill, L. and Allums, A. *A Dyadic Approach to the Detection of Deception*. Paper presented at the The 44th annual Hawaii International Conference on Systems Sciences, Koloa, HI, 2011.

[7] Dunbar, N. E., Jensen, M. L., Kelley, K. M., Robertson, K. J., Bernard, D. R., Adame, B. and Burgoon, J. K. *Cheating and Credibility: How modality and power affect veracity detection*. Paper presented at the Annual meeting of the National Communication Association, San Francisco, CA, 2010.

[8] Feeley, T. H. and deTurck, M. A. The behavioral correlates of sanctioned and unsanctioned deceptive communication. *Journal of Nonverbal Behavior, 22*(3), 189-204, 1998.

[9] Giordano, G. A., Stoner, J. S., Brouer, R. L., & George, J. F. The Influences of Deception and Computer-Mediation on Dyadic Negotiations. *Journal of Computer-Mediated Communication, 12*(2), 362-383, 2007.

[10] Goleman, D. *Social intelligence: The new science of human relationships.* New York: Random House, 2006.

[11] Grice, H. P. *Studies in the ways of words*. Cambridge: Harvard University Press, 1989.

[12] Jensen, M. L., Lowry, P. and Jenkins, J. Effects of Automated and Participative Decision Support in Computer-Aided Credibility Assessment. *Journal of Management Information Systems, 28*(1), 201-233, 2011.

[13] Kassin, S. M., Leo, R. A., Meissner, C. A., Richman, K. D., Colwell, L. H., Leach, A.-M. and La Fon, D. Police interviewing and interrogation: A self-report survey of police practices and beliefs. *Law and Human Behavior, 31*(4), 381-400, 2007.

[14] LaFrance, M. Postural mirroring and intergroup relations. *Personality and Social Psychology Bulletin, 11*(2), 207-217, 1985.

[15] Levenson, R. W. and Ruef, a. Physiological aspects of emotional knowledge and rapport. In W. Ickes (Ed.), *empathic accuracy* (pp. 44-72). New York: Guilford, 1997.

[16] Navarro, J. A Four-Domain Model for Detecting Deception. *FBI Law Enforcement Bulletin, 72*(6), 19, 2003.

[17] Turvey, B. E. Introduction to terrorism: Understanding and interviewing terrorists *Criminal profiling: An introduction to behavioral evidence analysis (3rd ed.).* (pp. 523-538). San Diego, CA US: Elsevier Academic Press, 2008.

[18] Vrij, A. Why professionals fail to catch liars and how they can improve. *Legal and Criminological Psychology, 9*, 159-181, 2004.

[19] Vrij, A. *Detecting lies and deceit: Pitfalls and opportunities* (2nd ed.). U.K.: Wiley, 2008.

[20] White, C. H. and Burgoon, J. K. Adaptation and communicative design: Patterns of interaction in truthful and deceptive conversations. *Human Communication Research, 27*(1), 9-37, 2001.

[21] A. Kanaujia, Y. Huang, and D. N. Metaxas. Tracking facial features using mixture of point distribution models. In *ICVGIP*, pages 492–503, 2006.

[22] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models-their training and application. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[23] S. Zhang, Y. Zhan, M. Dewan, J. Huang, D. Metaxas, and X. Zhou. Sparse shape composition: A new framework for shape prior modeling. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1025–1032, 2011.

[24] S. Zhang, Y. Zhan, M. Dewan, J. Huang, D. Metaxas, and X. Zhou. Towards robust and effective shape modeling: Sparse shape composition. *Medical Image Analysis*, 16(1):265–277, 2012.

[25] P. Yang, Q. Liu, and D. N. Metaxas. Rankboost with l1 regularization for facial expression recognition and intensity estimation. In *ICCV*, pages 1018–1025, 2009.

[26] J. K. Burgoon, B. A. L. Poire, and R. Rosenthal. Effects of preinteracton expectancies and target communication on perceiver reciprocity and compensation in dyadic interaction. *Journal of experimental social psychology*, pages 287–321, 1995.

[27] D. Metaxas. Physics-based deformable models: applications to computer vision, graphics, and medical imaging, volume 389. Springer, 1997.

[28] D. Decarlo and D. Metaxas. Optical flow con- straints on deformable models with applications to face tracking. *International Journal of Computer Vision*, 38(2):99–127, 2000.

[29] F. Yang, J. Wang, E. Shechtman, L. Bourdev, and D. Metaxas. Expression flow for 3d-aware face component transfer. *ACM Trans. Graphics (Proc. SIGGRAPH)*, 27(3):60, 2011.