Say CHEESE: Common Human Emotional Expression Set Encoder and its Application to Analyze Deceptive Communication

Taylan Sen, Md Kamrul Hasan, Minh Tran, Matthew Levin, Yiming Yang, and Mohammed (Ehsan) Hoque
Department of Computer Science
University of Rochester
Rochester, NY, United States
[tsen, mhasan8, mehoque]@cs.rochester.edu, [mtran14, mlevin6, yyang70]@ur.rochester.edu

***Abstract*— In this paper we introduce the Common Human Emotional Expression Set Encoder (CHEESE) framework for objectively determining which, if any, subsets of the facial action units associated with smiling are well represented by a small finite set of clusters according to an information theoretic metric. Smile-related AUs (6,7,10,12,14) in over 1.3M frames of facial expressions from 151 pairs of individuals playing a communication game involving deception were analyzed with CHEESE. The combination of AU6 (cheek raiser) and AU12 (lip corner puller) are shown to cluster well into five different types of expression. Liars showed high intensity AU6 and AU12 more often compared to honest speakers. Additionally, interrogators were found to express a higher frequency of low intensity AU6 with high intensity AU12 (i.e. polite smiles) when they were being lied to, suggesting that deception analysis should be done in consideration of both the message sender's and the receiver's facial expressions.**

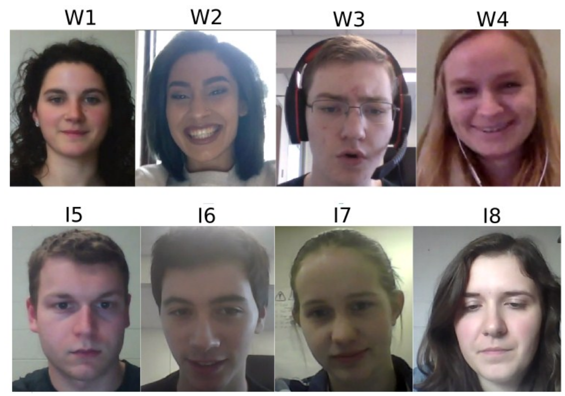

Figure 1. Witness and Interrogator Faces During Honest and Dishonest Communication *(Witness faces: W1-W4; Interrogator faces I5-I8). Answers are provided in the acknowledgement section*.

I. Introduction

Of the witnesses shown in Fig. 1 (W1-W4), can you guess who is being honest and who is lying? What about the interrogators (I5-I8) - can you tell which of them are being lied to and which are being told the truth? Could their smiles be giving us any clues?

Smiles are one of the most ubiquitously expressed facial expressions [1]. While common wisdom suggests that facial expressions associated with basic emotions are universal, recent research has suggested otherwise [2][3][4]. Smiles have been shown to have vastly different meanings, with variation in associated emotions spanning across opposite ends of the valence spectrum including happiness, frustration, embarrassment, and amusement [4][5][24][28][29][39].

While comprehension of smiles has been confounded by their multiple possible meanings, some promise has been shown in identifying specific movement patterns for several different meanings of a smile [4][22][23]. The Facial Action Coding System (FACS) has been widely used as a way of concretely defining and measuring specific movements comprising a facial expression in a given face [7]. In FACS, expressions are broken down into constituent levels of Facial Action Units ("AU1-AU45"), which independently correspond to facial movements roughly associated with individual muscle groups. Different types of smile expressions were studied as early as 1862 by Guillaume Duchenne, who identified that a smile marked by expression in both the regions around the eyes (AU6 & AU7) and corners of the mouth (AU12) is associated with spontaneous and sincere emotions of happiness, pleasure, and delight [6]. (Such a smile is now commonly referred to as a *Duchenne smile.*) Alternatively, consciously produced "polite" smiles were found by Duchenne to typically involve only the corners of the mouth without any contemporaneous contraction of muscles around the eyes. Smiles, and more generally facial expressions, can be a product of both conscious and unconscious behavior. By understanding unconscious smile expressions, we may gain insight to an individual's affective or mental state, possibly even when they are trying to deceive us. Following Duchenne's footsteps, much research has focused on identifying the differences between conscious (posed) and unconscious (spontaneous) smiles [22, 23]. In addition to researching the differences in how subjects felt while smiling, investigators have also examined how different smiles are perceived. Ambadar, et al. found that in comparison with smiles that are perceived as polite, smiles that are perceived as amused more often include the eyelid contraction and mouth opening. Schmidt, et al. found that lip corner movement asymmetry was not greater in deliberate smiles [33]. Frank et al. identified five separate markers to distinguish enjoyment and non-enjoyment smiles [34].

In addition to associating individual movements to spontaneous and posed smiles, researchers have also attempted to identify more diverse categories of smile types. Ekman, et al., has suggested that there are three types of smile including genuine, false, and miserable [24]. Ambadar et al., examined three slightly different categories of smiles including amused, polite, and embarrassed/nervous. Harris, et al. divided smiles into four types including the Duchenne smile, a "nonDuchenne smile" characterized by lip corner puller contraction only, a "controlled smile" marked by lip corner puller together

978-1-5386-2335-0/18/$31.00 ©2018 IEEE

with another lower face masking movement, and a "mixed smile" in which lip corner puller occurs together with a movement which has been associated with negative affect (AU1+4, AU10, and AU20). Each of the above studies relies crucially upon either knowing a participant's felt emotion, or the manual labeling of a perceived emotion by an independent judge. A study's confidence in a participant's reported or induced emotion could be questionable. Self reported surveys may not be accurate as participants often have no incentive to answer truthfully, or may be uncertain, unaware, or unable to articulate the true emotion they are feeling. Studies which have the subject perform an action which is typically associated with a positive or negative feeling might not happen to work with a given individual. It is typically costly and difficult to obtain large data sets of labels on smile pictures or recordings making many advanced machine learning approaches difficult to apply. Additionally, many past categorizations of smile types may also be inherently biased with our own psychological preconceptions, with categories being based on our imperfect definitions of emotions. Is there an objective way to determine how many types of smile there are from millions of unlabeled faces? By identifying such a set of common smiles can we gain insight into more complex behavior, such as deception?

Facial expressions pertaining to trustworthiness, negotiation, and deception have also been heavily studied [30][31][32][35][36][37][40]. Research has shown that smiles have been associated with deception under a psychological theory called *duping delight* [11][12]. Duping delight is the premise that deceivers take delight in lying to another, especially when there is an audience to the deception [11]. However, statistical analyses of individual differences in automatically extracted AU6 and AU12 between deceivers and honest speakers have found mixed results in their ability distinguish deceptive and honest communication [13][14]. By analyzing deception in the context of a smile model containing a finite number of objectively determined smile types, can we improve our understanding of deceptive facial expressions?

In summary, in this paper we show that by clustering a corpus of 1.3M frames by the right AU subsets (AU6 and AU12) and the right number of clusters (five) to optimize an information theoretic metric (CH score) we find that:

- Liars use strong Duchenne smile (high intensity AU6 & high intensity AU12) expressions significantly more often than honest speakers supporting the psychological theory of duping delight.
- Honest speakers use significantly more cheek raiser only (high intensity AU6, low intensity AU12) faces than liars.
- Even though interrogators were unable to identify liars more often than random chance, the interrogators' faces showed significantly higher levels of low intensity AU6 high intensity AU12 faces when their witnesses were lying versus telling the truth.

II. Methods

In this section, we first provide some details about the dataset and the tools used to extract raw facial expression into a multidimensional set of components. Then we describe the clustering methodology used with emphasis on selecting the cluster number and dimensional subsets. This section then finally covers the statistical tools we used in the analysis.

A. Dataset

In this study we used video recordings of 151 pairs of individuals playing a communication game involving deception using the ADDR framework as described in our prior work [14]. In summary, both university students and Amazon Mechanical Turkers [15] were recruited to play a five-minute deception game with another participant over the Internet using a videochat-enabled web application in return for $10. (The total time involving directions training and reading the IRB waiver took ~30 minutes.) After two participants were video-linked, the web application randomly assigned one participant the role of witness and the other participant the role of interrogator. The witness was shown evidence (i.e. an image) by the web application for 30 seconds and instructed to memorize its details. After 30 seconds, the web application randomly instructed the witness to either lie (i.e. a sanctioned lie), or tell the truth regarding the image to the interrogator for the rest of the interaction. The interrogator was instructed by the web application to question the witness about the evidence in order to determine if the witness was lying or telling the truth. Additionally, the interrogator was provided specific questions to ask the witness by the web application. The interrogator was also encouraged to ask their own questions. At the end of five minutes of questioning, interrogators were asked by the web application whether they thought the witness was lying or telling the truth. Interrogators received a bonus ($5-$10) if they correctly determined whether the witness was lying or telling the truth regarding the image. The witness received a bonus ($5-$10) if the interrogator thought he/she was telling the truth (regardless of whether the witness was indeed lying or telling the truth). The quality of the data was controlled by ensuring the bandwidth and video quality of the participants' computer systems. Additionally, participants were required to pass a test on their understanding of the game protocol before they were allowed to participate.

The questions that the interrogators were directed to ask the witnesses involved two different phases. First, a set of baseline questions, which had nothing to do with the evidence image were asked. The baseline questions were fully scripted in that the interrogator was not given any leeway in what they were supposed to ask. Next, the interrogator was directed to ask questions relevant to identifying the evidence ("relevant questions"). The relevant questions were also a number of scripted questions provided to the interrogator. Following the relevant questions, the web application then directed the interrogator to then ask the witness his/her own questions.

In this dataset, the ground truth of whether the witness is lying or telling the truth was known with near absolute certainty since the web application kept track of the witness role assignment. Additionally, videos were manually reviewed to ensure participants were following directions. The video recordings of each participant were all webcam based head front-facing videos recorded at 15 fps.

The interrogators in this study were not professional interrogators, but participants from the same pool as the witnesses. Both interrogators and witnesses were not screened for their skill at telling or detecting lies, as our aim is to study members of the general population, not trained professionals or highly experienced individuals. Due to the automated nature of running the protocol through the web application, witnesses were shown evidence in an identical manner. Similarly, interrogators were directed to question their witnesses in an identical way. Thus, the experimental context was highly preserved among the study data, leading to a standardized dataset.

B. Facial Analysis

The video recordings were analyzed with the open source OpenFace analysis tool [8]. For each frame of video, OpenFace outputs a set of 17 facial action unit (AU) outputs. For each of the action units extracted, OpenFace provides both a Boolean output, indicating presence or absence of facial unit expression, as well as a continuous, intensity level output. In this study, we only used the continuous outputs from OpenFace. The intensity level output is a continuous output in the range of [0,5], zero representing no expression, and 5 representing maximal expression.

OpenFace has been shown to provide an accurate measure of action units that is supported by peer reviewed benchmarks [8]. Due to its automated nature, we expect OpenFace to be fairly objective. Frames with an OpenFace tracking confidence below 90% were excluded from analysis.

C. Clustering

In order to identify whether facial expressions expressed during dyadic communication form clusters among facial action unit subsets, we used the k-Means algorithm [16]. K-Means is an iterative algorithm to find the set of k cluster centers which minimizes the squared distance between the points in a given data set and each data point's closest cluster center. More specifically, we model each frame of each video as a separate face datapoint defined by a vector of AU levels. Various subsets of AUs were explored for a range of different cluster numbers k. Because the k-means algorithm is not guaranteed to find the optimal solution (since it may get stuck in local optima), we re-ran each attempt at clustering for a given set of AUs and k 10 times each with different random initializations, and used the best performing solution.

D. Identifying the Ideal Number of Clusters

The k-Means algorithm requires the number of clusters (k) to be provided. Because the average error in representing a corpus of faces with a finite number of clusters will always go down when using more clusters, we must use another way of determining an ideal number of clusters. The Calinski Harabasz criterion ("CH score") is a well-known method for selecting the number of clusters based on ANOVA ideology [17]. The CH score measures cluster quality based on a comparison of the differences of datapoints within a cluster to the differences of datapoints between clusters. Therefore, a maximized CH score finds a low cluster number in which clusters have low variance and high inter-cluster distance. The CH score has been shown to perform well in both experimental and theoretical clustering experiments. Milligan and Cooper [18] showed that the CH score outperformed 29 other k-search "stopping rules" when clustering with multidimensional data. In another comparative study Olatz et al. [19] classified cluster validity indices into three groups based on their performance, in which CH score is among the best performing group. Effectively, the CH score provides an objective measure to determine an ideal small number of smile types to describe a large number of faces (1.3M).

E. Subset Selection

Clustering in high dimensional space has been shown to be problematic [25]. In a high dimensional space with a number of noisy dimensions, the Euclidian distance between two points is likely to be dominated by noise. Thus, if multiple AUs are noisy, we should not expect CH score to identify a good k-means clustering result. Indeed, when running k-means analysis with all 17 action units provided by OpenFace, the CH score showed a continually decreasing score as k was increased from 2 through 14 (Fig. 2). No meaningful clusters based on CH score were identified when using all AUs. Because good clustering may exist when noisy dimensions are removed, we needed to identify the important subsets from 17 action units that produce good clusters. However, the power set of all action units is large and it is not feasible to examine all subsets. (For example, when searching with k = 2 to 14, with 2^{17} subsets this would entail > 1,000,000 runs of the k-means algorithm.) We reduced the search space by considering the action units that are commonly used during smiling. Previous study showed that people smile frequently during social interactions and that it increases trust among strangers [38]. Actions units related to smile can be key factors to cluster common facial expressions that people use during communication. In order to determine which AU subsets produce a good clustering result, we conducted a brute-force search through all AUs related to smile including: AU6 (cheek raiser), AU7 (lid tightener), AU12 (lip corner puller), and AU14 (dimpler).

F. Statistical Analysis

In order to compare the frequency of clusters observed in deceptive vs. honest witnesses, we use the Mann-Whitney U Test (aka "ranksum test") [20]. The U test is used to compare the medians of two random variables. Unlike the popular Student's t-test, it does not rely on the assumption that the populations being compared are normally distributed. This is of particular concern to us since the distributions of AU intensity levels is bounded to the interval [0,5], unlike a normal distribution which has non-zero probability density across [-∞,∞]. It was expected and experimentally verified that the frequency of neutral faces (i.e. AU intensities ~0) was substantially greater than other faces, making the U test particularly appropriate.

III. Results and Discussion

Each of the 302 video files from 151 individuals either playing the role of interrogators or witnesses were

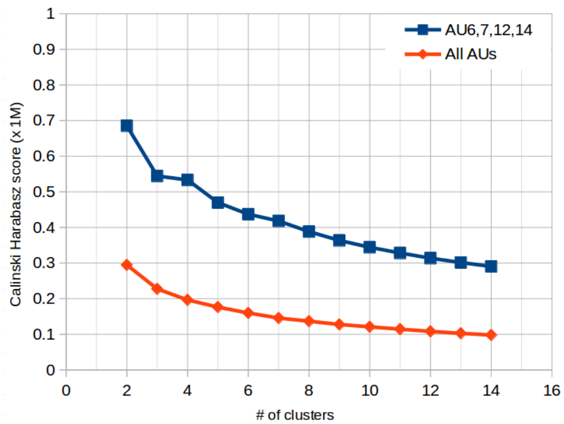

Figure 2. Calinski Harabasz score for all AUs and smile-related AUs.

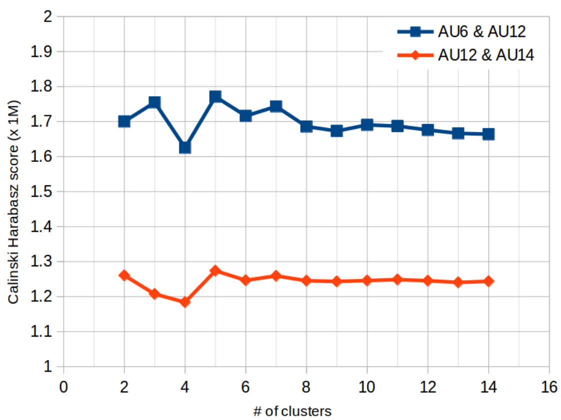

Figure 3. Calinski Harabasz score for subsets showing an ideal number between 2 and 15.

processed with OpenFace. In the following subsections we describe the results and analysis of clustering all AUs, smile-related AUs, and all subset pairs of smile-related AUs for differing cluster numbers.

A. Searching for the ideal number of clusters k

K-Means clustering was run using k equal to all integer values from 2 to 14. Recall that for each the k-means is run 10 times with different random initializations in order to increase the chances that we have a run which finds the global optimal solution without getting stuck in a local minima. The CH scores for clustering with all of the intensity-based AUs (1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 20, 23, 25, 26, 45) provided by OpenFace as well as with only smile related AUs (6, 7, 12, 14) are shown in Fig. 2. For both of these sets of AUs, the best CH score was found with k = 2. Due to the large proportion of neutral faces, the clustering results always included a cluster centered near the origin. Thus, results with k=2 are trivial, indicating a failure to identify a meaningful clustering result.

Out of all subset pairs of the smile-related AUs, the only subsets with a resulting ideal cluster number k between 2 and 15 are (AU6 with AU12) and (AU12 with AU14) as shown in Fig. 3. As shown in Fig. 3, the k value

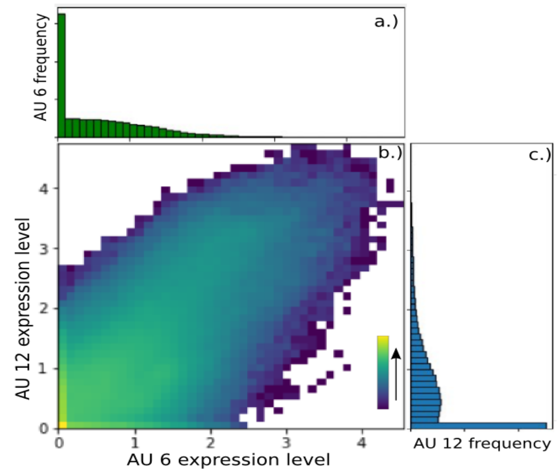

Figure 4. Distribution of AU6 and AU12 expression levels across all 1.3M faces. a.) histogram of AU6 expressions levels alone, b.) 2D histogram of AU6 and AU12 simultaneously, c.) histogram of AU12 expression levels alone.

which maximizes the CH score for both of these AU subsets is 5.

Shown in Fig. 4a-c are the histograms of the AU6 and AU12 levels for each observed face in the dataset. It is important to note in Fig. 4a-c the large proportion of faces which have near-zero value for AU6 or AU12. This is observed as the large spikes near the origin in Fig. 4a and Fig. 4c as well as the bright yellow bin in the bottom left corner of Fig. 4b. The high frequency of faces that have a zero value for only either AU6 or AU12 is also demonstrated in Fig. 4b by the bright horizontal and vertical ridges along the x and y axes. Additionally, in Fig. 4c, we can see that there is a frequency peak at AU12 level of around 0.5. We should not be surprised if the ideal clustering solution contains cluster centers associated with one or more of these high frequency regions. Fig. 4b also demonstrates the positive correlation between AU6 and AU12. More often than not, it appears that when AU6 is expressed, so is AU12. This observation is further supported by the Pearson correlation between AU6 and AU12 r=0.70.

B. Clustering Results for All Faces with AU6 and AU12

Each of the face datapoints as well as the cluster centers resulting from running k-Means with k=5 are shown in Fig. 5, and the AU6 and AU12 values associated with each of the identified cluster centers are displayed in Table I. In addition, for each cluster a representative face has been added from the associated video files for which the extracted AU levels matched the cluster. As expected from our histogram analysis in Fig. 4, we find that one of the cluster centers is located near the AU6 = 0 AU12 = 0 region, which is labeled as cluster 0, "Neutral Face". As shown in Table I, the AU6 and AU12 values for cluster 0 are 0.11 and 0.15 respectively, which are indeed are close to 0 (considering the AUs are measured on a 0-5 scale). Cluster 1, has AU6 and AU12 values of 0.34 and 0.95 respectively. Thus, Cluster 1 appears to match the definition of a non-Duchenne smile since it contains

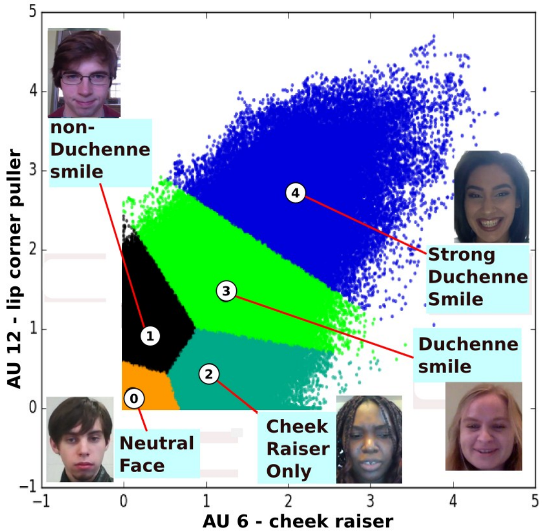

Figure 5. Results of clustering based on AU6 and AU12 across all faces.

substantial AU12 expression with a significantly lower level of AU6 expression. Cluster 2 represents faces which have mostly AU6 expression with little AU12 and are thus labeled as "Cheek Raiser Only". Clusters 3 and 4 represent faces in which AU6 roughly equals AU12 expression, at moderate and high levels respectively. These clusters are thus labeled as cluster 3 "Duchenne smile" and cluster 4 "Strong Duchenne Smile".

The clustering shown in Fig. 5 was conducted using all frames (in which OpenFace had a tracking confidence greater than 90%) of all video files, including both interrogators and witnesses, in situations where both the witness was lying and telling the truth. Clustering with k=5 was repeated using different subsets of the faces, each producing results which largely resembled the results in Fig. 5 and Table I.

TABLE I. CLUSTER CENTERS DEFINED BY K-MEANS WITH K=5

Face Cluster	Cluster Centers *AU6*	Cluster Centers *AU 12*
0 – Neutral Face	0.097	0.121
1 – non-Duchenne Smile	0.31	0.90
2 – Cheek Raiser Only	1.03	0.47
3 – Duchenne Smile	1.23	1.51
4 – Strong Duchenne Smile	2.09	2.72

C. *Distribution of the face clusters among witnesses and interrogators in honest and dishonest communication*

We next examined whether the proportion of each face cluster expressed by witnesses and interrogators differed based on whether the communication was honest or deceptive. The cluster centers provided from k-means, displayed in Fig. 5 and Table I, were used to develop an encoder which encodes a face represented by an AU6 and AU12 value into a cluster number.

Because the baseline questions were not supposed to be involved in any deception, they were not included in the analysis. This encoder was thus applied to the video frames from the relevant questioning phase only. The cluster distribution data results for witnesses and interrogators are shown in Table II and Table III respectively. In addition to providing the percentage of each face cluster that shows up in a group's total faces, Tables II and III also show the Mann Whitney ranksum test p-values and Cohen's d effect size for comparing deceptive and honest communication.

As shown in Table II, both honest and deceptive witnesses had a large percentage of their faces encoded as neutral faces (43.3 and 41.77% respectively). The Mann Whitney U test did not show any significant difference between the percentage of neutral faces in honest and deceptive witnesses. The percentage of non-Duchenne faces expressed by honest and dishonest witnesses were also similar at 22.66% and 24.28%. However, the percentage of Cheek Raiser face clusters between honest and dishonest witnesses was significantly different with truthful witnesses expressing a Cheek Raiser Only face 19.30% of the time compared to 15.40% for dishonest witnesses. The ranksum test p-value is 0.01 and Cohen's effect size d=0.16 showing a small effect size. While the percentage of Duchenne Smile (cluster 3) faces was not significantly different between honest and dishonest witnesses, the percentage of Strong Duchenne Smile faces (cluster 4) did show significant difference. Dishonest witnesses were shown to express Strong Duchenne smiles 5.50% of the time compared to only 3.30% for honest witnesses.

The differences in Strong Duchenne smile expression between honest and dishonest witnesses is consistent with the theory of duping delight. More specifically, since Duchenne smiles are associated with pleasure, if deceivers are indeed having a good time telling a lie as predicted by duping delight, we would expect to see a higher frequency of Duchenne smiles in deceivers. Since the lying witnesses know that the researchers will ultimately know that they were lying to the interrogators, the researchers are effectively acting as an audience. An audience has been described to amplify the duping delight effect [11]. It should be noted the frequency of moderate Duchene smile

TABLE II. DISTRIBUTION OF FACE TYPES IN WITNESSES

Face Cluster	Cluster Distribution Profile: *Truthful Witnesses*	Cluster Distribution Profile: *Dishonest Witnesses*	Mann Whitney ranksum test (p-value)	Cohen's d effect size
0 – Neutral Face	43.3%	41.8%	0.425	0.046
1 – non-Duchenne Smile	22.7%	24.3%	0.397	-0.069
2 – Cheek Raiser Only	**19.3%**	**15.4%**	**0.01**	**0.164**
3 – Duchenne Smile	11.5%	13.0%	0.200	-0.106
4 –Strong Duchenne Smile	**3.3%**	**5.5%**	**0.033**	**-0.258**
All clusters	100.0%	100.0%		

TABLE III. DISTRIBUTION OF FACE TYPES IN INTERROGATORS

Face Cluster	Cluster Distribution Profile: *Interrogator of Truthful Witnesses*	*Interrogator of Dishonest Witnesses*	Mann Whitney ranksum test	Cohen's d effect size
0 – Neutral Face	44.60%	36.20%	0.101	0.25
1 – non-Duchenne Smile	**18.30%**	**26.50%**	**0.031**	**-0.361**
2 – Cheek Raiser Only	21.20%	18.90%	0.097	0.085
3 – Duchenne Smile	12.40%	14.00%	0.125	-0.097
4 – Strong Duchenne Smile	3.60%	4.40%	0.211	-0.120
All clusters	100.00%	100.00%		

(cluster 3) is also higher in deceptive witnesses (13.01% compared to 11.45%), however the difference does not reach statistical significance.

An interesting finding is that although the percentage of Duchenne smile face clusters shows a statistically significant difference between honest and dishonest witnesses, when independently looking at average AU6 and AU12 levels between honest and dishonest witnesses, there is no statistically significant difference [14]. It is only after encoding the faces into clusters that a statistically significant difference is observed.

We were unable to identify prior work which explains why higher Cheek Raiser Only faces would be expressed in truthful witnesses. An ad hoc analysis of truthful witnesses expressing Cheek Raiser Only faces suggested that this face cluster is an expression consistent with the act of trying to remember what the evidentiary image was. The woman shown in the Cheek Raiser Only cluster in Fig. 5 indeed was in the middle of recollecting and describing details from the image. However, a more thorough study needs to be conducted in order to test this hypothesis.

While the Duchenne smile frequencies are reasonably explained by duping delight, an alternative explanation is that deceptive witnesses are trying to get the interrogators to like or trust them. Smiles have been associated with trust [38]. By inducing interrogators to like or trust them, witnesses may be making it harder for the interrogator to accuse them of lying. Since a lying witness's bonus depends on fooling the interrogator, they have a directly financial incentive to mislead the interrogator. In order to indirectly evaluate this conjecture, we look to the interrogators' facial expressions.

As shown in Table III, interrogators showed significant differences in their face cluster distributions only in their expression of non-Duchenne smiles (cluster 1). More specifically, interrogators who were paired with dishonest witnesses demonstrated a marked increase in their percentage of non-Duchenne smiles (26.5% compared to 18.30% for interrogators paired with honest witnesses). The p-value for this difference was 0.031, and Cohen's d showed a moderate effect size at d=0.36.

While it may be unintuitive to expect interrogators to express different facial expressions when they are unaware of the witnesses honesty, it is consistent with the communications theory Interpersonal Deception Theory ("IDT") [21]. IDT studies have demonstrated that communication can be an interactive event between a message sender and a message receiver which is disrupted when there is deception. A lack of mirroring (i.e. interrogators display of a polite smile instead of mirroring the witness's Duchenne smile) could be viewed as a disruption of the synchrony between message senders and receivers. However, while this may explain a disparity between witnesses and interrogators, it does not explain why the interrogators specifically are displaying higher levels of non-Duchenne smiles in particular.

Since the interrogators were correct in detecting that the witness was lying only 48% of the time it is unlikely that the interrogators' higher non-Duchenne smile rate is due to an amused smile due to catching the witness in a lie. It is worth noting that the interrogators of truthful witnesses had a higher percentage of Neutral Faces (44.6% compared to 36.2%). While this difference is not statistically significant according the Mann Whitney ranksum test p-value, the Cohen's d represents a reasonable difference of d=0.25. The fact that interrogators of truthful witnesses display more neutral faces is consistent with the concept that lying witnesses are trying to befriend, entertain, and/or cajole the interrogators into believing them.

One limitation of this research is that the clustering algorithm does not take advantage of the correlations that exist between particular AUs. A clustering algorithm which uses Mahalanobis distance, preprocesses the AUs with PCA, or a mixture model of distributions with non-zero correlation could be used to theoretically better fit the distribution of faces.

It should also be noted that the findings on smiles regarding deception were obtained from a protocol involving sanctioned lies (i.e. lies which the participants were instructed to tell.) It is thus possible that the lying witnesses did not experience the same emotions and feelings that an individual who decided to lie on their own would. However, past research has found that there is no perceivable differences in the behavior of sanctioned versus unsanctioned liars [26]. Other research, however, has identified that there are differences between the two groups [27]. Whether there is a difference or not, it is still worthwhile to study the behavior of sanctioned liars. Indeed if we allow participants to decide whether to lie or not, we will no longer have an unbiased distribution of deceptive and honest witnesses. If participants could choose whether to lie, perhaps only participants who are good at or comfortable with lying would decide to lie. By randomly assigning whether participants need to lie or tell the truth, we are able to maintain unbiased distributions of participants. In our future work, we will run a similar protocol in which participants will be allowed to decide on their own whether to lie or not to quantify any observable differences between the two groups if any.

Additionally, our protocol involved a game-like setting which may affect the context of lies and truths told. While this may elicit more feelings of fun and reduce feelings of guilt, it may also act to increase the stakes for competitive participants. Further, the game-like setting may induce stronger effect of duping delight. Regardless whether the protocol was a game-like setting, the protocol involved real financial compensation that depended upon deceitful versus honest behavior. A deceitful witness was given an additional bonus if they successfully bluffed without being detected. This reward, and the risk of loss from not receiving it, may simulate natural instances of deception in real-life scenarios by simulating the potential gain or loss that may be associated with lying. Additional guilt may be present for the bluffing witness since they are preventing the interrogator from receiving their financial bonus if they successfully lie without the interrogator detecting deceit.

It is important to note that differences may exist between telecommunication based deception and in-person deception. However, due to the increasingly ubiquitous nature of webcam-based telecommunication, it is essential to understand deceit via webcam in its own right.

In our future work, we plan to exhaustively investigate subsets of all AUs as well as alternative clustering algorithms. The dataset used in this study will be made public within the next year.

IV. Conclusion

Communication, and especially communication through facial expressions, is a very context dependent phenomena. It is likely that findings with one set of communicators following a particular protocol might not hold with another group or a different protocol. We have demonstrated a methodological framework for automatically identifying a set of smile types through unsupervised learning and applied this framework to a corpus of dyadic communication. We showed that in this corpus, the only smile related AUs which cluster well are AU6 and AU12, largely vindicating the past historical research focus on these two AUs. This framework can be readily applied to other face corpora as differing contexts become available. Additionally, the value of developing an encoder based upon this framework was demonstrated by identifying differences between honest and dishonest communication which were not apparent when using raw AUs alone. Additionally, the methodology described in this paper can be easily extended to search the space of AUs other than just those associated with smile.

Acknowledgment

This work was supported by Grant W911NF-15-1-0157 with the Army Research Office (ARO). The authors would also like to express their gratitude to the ROC-HCI group for their many helpful comments and help in improving this paper. Regarding Fig. 1, witnesses W2 and W4 were lying. Interrogators I6 and I7 were interrogating a lying witness.

References

[1] R. W. Buck, V. J. Savin, R. E. Miller, and W. F. Caul, "Communication of affect through facial expressions in humans." Journal of Personality and Social Psychology, vol. 23, no. 3, p. 362, 1972.

[2] M. Gendron, D. Roberson, J. M. van der Vyver, and L. F. Barrett, "Perceptions of emotion from facial expressions are not culturally universal: evidence from a remote culture." Emotion, vol. 14, no. 2, p. 251, 2014.

[3] R. E. Jack, C. Blais, C. Scheepers, P. G. Schyns, and R. Caldara, "Cultural confusions show that facial expressions are not universal," Current Biology, vol. 19, no. 18, pp. 1543– 1548, 2009.

[4] M. E. Hoque, D. J. McDuff, and R. W. Picard, "Exploring temporal patterns in classifying frustrated and delighted smiles," IEEE Transactions on Affective Computing, vol. 3, no. 3, pp. 323–334, 2012.

[5] M. Hoque and R. W. Picard, "Acted vs. natural frustration and delight: Many people smile in natural frustration," in Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on. IEEE, 2011, pp. 354–359.

[6] P. Ekman, R. J. Davidson, and W. V. Friesen, "The duchenne smile: Emotional expression and brain physiology: Ii." Journal of personality and social psychology, vol. 58, no. 2, p. 342, 1990.

[7] P. Ekman and W. V. Friesen, Facial Action Coding System: Investigatoris Guide. Consulting Psychologists Press, 1978.

[8] T. Baltrušaitis, P. Robinson, and L.-P. Morency, "Openface: an ˇ open source facial behavior analysis toolkit," in Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on. IEEE, 2016, pp. 1–10.

[9] G. Littlewort, J. Whitehill, T. Wu, I. Fasel, M. Frank, J. Movellan, and M. Bartlett, "The computer expression recognition toolbox (cert)," in Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on. IEEE, 2011, pp. 298–305.

[10] D. McDuff, A. Mahmoud, M. Mavadati, M. Amr, J. Turcot, and R. e. Kaliouby, "Affdex sdk: a cross-platform real-time multi-face expression recognition toolkit," in Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems. ACM, 2016, pp. 3723–3726.

[11] P. Ekman, Telling lies: Clues to deceit in the marketplace, politics, and marriage (revised edition). WW Norton & Company, 2009.

[12] L. Ten Brinke and S. Porter, "Cry me a river: Identifying the behavioral consequences of extremely high-stakes interpersonal deception." Law and Human Behavior, vol. 36, no. 6, p. 469, 2012.

[13] M. Lewis, C. Stanger, and M. W. Sullivan, "Deception in 3- year-olds." Developmental psychology, vol. 25, no. 3, p. 439, 1989.

[14] T. Sen, M. K. Hasan, Z. Teicher, and M. E. Hoque, "Automated dyadic data recorder (ADDR) framework and analysis of facial cues in deceptive communication," Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 1, no. 4, p. 163, 2018.

[15] G. Paolacci, J. Chandler, and P. G. Ipeirotis, "Running experiments on amazon mechanical turk," 2010.

[16] S. Lloyd, "Least squares quantization in pcm," IEEE transactions on information theory, vol. 28, no. 2, pp. 129–137, 1982.

[17] T. Caliński and J. Harabasz, "A dendrite method for cluster ´ analysis," Communications in Statistics-theory and Methods, vol. 3, no. 1, pp. 1–27, 1974.

[18] G. W. Milligan and M. C. Cooper, "An examination of procedures for determining the number of clusters in a data set," Psychometrika, vol. 50, no. 2, pp. 159–179, 1985.

[19] O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J. M. PéRez, and ´ I. Perona, "An extensive comparative study of cluster validity indices," Pattern Recognition, vol. 46, no. 1, pp. 243–256, 2013.

[20] H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," The annals of mathematical statistics, pp. 50–60, 1947.

[21] D. B. Buller and J. K. Burgoon, "Interpersonal deception theory," Communication theory, vol. 6, no. 3, pp. 203–242, 1996.

[22] Z. Ambadar, J. F. Cohn, and L. I. Reed, "All smiles are not created equal: Morphology and timing of smiles perceived as amused, polite, and embarrassed/nervous," Journal of nonverbal behavior, vol. 33, no. 1, pp. 17–34, 2009.

[23] H. Dibeklioğlu, A. A. Salah, and T. Gevers, "Recognition of ˘ genuine smiles," IEEE Transactions on Multimedia, vol. 17, no. 3, pp. 279–294, 2015.

[24] P. Ekman and W. V. Friesen, "Felt, false, and miserable smiles," Journal of nonverbal behavior, vol. 6, no. 4, pp. 238– 252, 1982.

[25] M. Steinbach, L. Ertoz, and V. Kumar, "The challenges ¨ of clustering high dimensional data," in New directions in statistical physics. Springer, 2004, pp. 273–309.

[26] T. H. Feeley, "Exploring sanctioned and unsanctioned lies in interpersonal deception," Communication Research Reports, vol. 13, no. 2, pp. 164–173, 1996.

[27] T. H. Feeley et al., "The behavioral correlates of sanctioned and unsanctioned deceptive communication," Journal of Nonverbal Behavior, vol. 22, no. 3, pp. 189–204, 1998.

[28] C. Harris and N. Alvarado, "Facial expressions, smile types, and self-report during humour, tickle, and pain," Cognition & Emotion, vol. 19, no. 5, pp. 655–669, 2005.

[29] D. Keltner, "Signs of appeasement: Evidence for the distinct displays of embarrassment, amusement, and shame." Journal of personality and social psychology, vol. 68, no. 3, p. 441, 1995.

[30] E. Krumhuber, A. S. Manstead, D. Cosker, D. Marshall, P. L. Rosin, and A. Kappas, "Facial dynamics as indicators of trustworthiness and cooperative behavior." Emotion, vol. 7, no. 4, p. 730, 2007.

[31] G. Lucas, G. Stratou, S. Lieblich, and J. Gratch, "Trust me: multimodal signals of trustworthiness," in Proceedings of the 18th ACM International Conference on Multimodal Interaction. ACM, 2016, pp. 5–12.

[32] A. Todorov, "Evaluating faces on trustworthiness," Annals of the New York Academy of Sciences, vol. 1124, no. 1, pp. 208– 224, 2008.

[33] K. L. Schmidt, Z. Ambadar, J. F. Cohn, and L. I. Reed, "Movement differences between deliberate and spontaneous facial expressions: Zygomaticus major action in smiling," Journal of Nonverbal Behavior, vol. 30, no. 1, pp. 37–52, 2006.

[34] M. G. Frank, P. Ekman, and W. V. Friesen, "Behavioral markers and recognizability of the smile of enjoyment." Journal of personality and social psychology, vol. 64, no. 1, p. 83, 1993.

[35] P. Ekman and W. V. Friesen, "Nonverbal leakage and clues to deception," Psychiatry, vol. 32, no. 1, pp. 88–106, 1969.

[36] V. Pérez-Rosas, M. Abouelenien, R. Mihalcea, and M. Burzo, ´ "Deception detection using real-life trial data," in Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. ACM, 2015, pp. 59–66.

[37] L. Ten Brinke, S. MacDonald, S. Porter, and B. O'connor, "Crocodile tears: Facial, verbal and body language behaviours associated with genuine and fabricated remorse." Law and human behavior, vol. 36, no. 1, p. 51, 2012.

[38] J. P. Scharlemann, C. C. Eckel, A. Kacelnik, and R. K. Wilson, "The value of a smile: Game theory with a human face," Journal of Economic Psychology, vol. 22, no. 5, pp. 617–640, 2001.

[39] M. Hoque, L.-P. Morency, and R. W. Picard, "Are you friendly or just polite?–analysis of smiles in spontaneous face-to-face interactions," in International Conference on Affective Computing and Intelligent Interaction. Springer, 2011, pp. 135–144.

[40] R. Li, J. Curhan, and M. E. Hoque, "Understanding social interpersonal interaction via synchronization templates of facial events," 2018.