LIWC into the Eyes: Using Facial Features to Contextualize Linguistic Analysis in Multimodal Communication

Md Kamrul Hasan
Department of Computer Science
University of Rochester
Rochester, USA
mhasan8@cs.rochester.edu

Taylan Sen
Department of Computer Science
University of Rochester
Rochester, USA
tsen@cs.rochester.edu

Yiming Yang
Department of Computer Science
University of Rochester
Rochester, USA
yyang70@u.rochester.edu

Raiyan Abdul Baten
Dept. of Electrical and Computer Engineering
University of Rochester
Rochester, USA
rbaten@ur.rochester.edu

Kurtis Glenn Haut
Department of Computer Science
University of Rochester
Rochester, USA
khaut@u.rochester.edu

Mohammed (Ehsan) Hoque
Department of Computer Science
University of Rochester
Rochester, USA
mehoque@cs.rochester.edu

***Abstract*—This paper demonstrates that analyzing language patterns in light of their associated facial expressions elicits significant differences between deceptive and truthful communication. Facial Action Units (AU) were analyzed in video recordings (1.2M frames) of 151 dyadic conversations following an interrogation protocol, in which one of the participants is known to be either lying or telling the truth. Linguistic features were extracted from the transcripts using Linguistic Inquiry and Word Count (LIWC) dictionary. Our framework extracted facial-feature contexts automatically corresponding to high and low intensities of AU occurrences. This helped us dive deeper into answers corresponding to the video segments where the witnesses kept their eyes wide open (high intensity of AU05–upper lid raise). We found that in these segments, deceivers used significantly fewer 'Seeing', 'Perceptual' and 'Cognitive' words and their answers were significantly shorter than truth-tellers.**

***Index Terms*—deception, multimodal, Action Unit, cognitive**

I. Introduction

From security screening checkpoints to ticket counters, online meetings to one-on-one conversations—the ability to detect deception automatically has its applications wherever security is important. Airport or border security officials need to be constantly vigilant to ferret out security threats, and audio-visual surveillance technologies add value to their ability to monitor hundreds of thousands of people every day. People all over the world use online video communication tools for business or personal interactions—security can be vital in such interpersonal dealings as well, among many other applications of automatic detection of deception. Over the years, researchers have studied non-verbal and verbal cues to detect deception [1]–[7]. However, the cues themselves are not absolute markers and can depend on their *contexts* [8], [9], making context-agnostic detection models prone to error.

Ekman developed the concept of micro-expression, which can reveal true emotions [10]. Researchers used these micro-expressions to unmask the true emotions of deceivers [1], [11]. The Facial Action Coding System (FACS) was developed by Ekman in an attempt to identify the movements of a set of visually discernible facial muscles [12]. These objective measurements, called Action Units (AU), are related to various emotions [13]–[15] and mental states [16]–[18], and give a systematic way to study micro-expressions. Researchers have also explored other non-verbal cues such as head movements [4], metaphoric and rhythmic hand gestures [2], pupil dilation [3] and eye blinking rates [19] to detect deception. Deception is also shown to affect the liar's respiration rate [20] and systolic blood pressure [21], thus giving away useful cues.

As for verbal cues, linguistic patterns have been studied widely, among other markers for detecting deception. Bluffers appear to use fewer first person and more third person singular pronouns than truthful people [5]. They tend to use words associated with negative emotions, and avoid exclusion words [6]. They have also been found to struggle delivering spatial information in their lies [7]. Linguistic Inquiry and Word Count (LIWC) provides a way to identify psychological meanings associated with words [22]. Researchers have widely adopted LIWC for analyzing human communication behavior towards understanding personality styles [23], social media behavior [24] and deception [5], [25], [26].

Thus, various verbal and non-verbal cues have been reported in literature to correspond to a person's complex thinking process during deception. However, these cues are not free from noise and can depend strongly on the associated contexts, irrespective of how truthful someone is actually being. For

This research was supported in part by grant W911NF-15-1-0542 and W911NF-19-1-0029 with the US Defense Advanced Research Projects Agency (DARPA) and the Army Research Office (ARO).

978-1-7281-3888-6/19/$31.00 ©2019 IEEE

example, Hoque *et al.* [8] showed that a smile does not always represent delight or happiness, and can in fact be a reflection of frustration. Without taking into account the context in which the cues are observed, the noisy effects can thus corrupt the psychological reasonings behind deception markers, leading to confounding results. This motivates conducting context-aware analyses of the multimodal cues. In their experiment, Vrij *et al.* [27] explored deception in the context of maintaining eye contact. They showed that the instruction to maintain eye contact increased the ability to distinguish between truths and lies from chance level (control condition) to above chance level (eye contact condition). They also found that liars provide less auditory and spatial details in their responses. Despite the promise of context-based analysis, it has remained a challenge for computers to automatically establish meaningful contexts that can be used to analyze relevant cues.

In this paper, we propose a method for retrieving contextual understanding of deception from facial expressions, and analyzing linguistic cues based on the extracted facial contexts. The key intuition here is, a high intensity of a particular facial signal might set the context of underlying mental processes, and linguistic patterns during that time segment might contain amplified information that distinguishes truth-tellers from bluffers. We analyze video data from a dyadic interrogation protocol where the ground truth is known. Our framework goes beyond merely combining facial Action Units and LIWC features. Instead, the framework first uses Action Units to automatically partition the videos into a number of separate segments based on "high" or "low" Action Unit intensities. Then, within the video segments, linguistic behavior is analyzed with LIWC features. Our results show that in the segments of *highly intense display of AU05* (upper lid raised/eyes wide open), fewer words from the 'Seeing', 'Perceptual' and 'Cognitive' LIWC categories are used by the deceivers. In addition, deceivers tend to provide shorter answers in those regions, compared to truth-tellers. This multi-step procedure, utilizing the facial Action Unit segmented contexts with LIWC features, is crucial to observing statistically significant results, and was not found in any of the identified prior art references. The closest references we could identify [28]–[32] do not involve any form of multi-step temporal segmenting and simply combines linguistic and facial expression features.

Our major contributions include:

- Developing a framework for automatically identifying conversation segments based on facial expression contexts, which affect language usage patterns in deceptive communication.
- Uncovering evidence that in the context of *highly intense display of AU05* (Upper Lid Raise), fewer 'Seeing', 'Perceptual' and 'Cognitive' words are used by deceivers, and their answers are shorter compared to truth-tellers.

II. Dataset

We use an interrogation game video dataset ($N = 151$ dyads) collected through the Automated Dyadic Data Recorder (ADDR) framework [33]. This dataset involves paired crowd-sourced participants playing a communication game, in which one was assigned to be an interrogator and the other a witness. At the beginning of the game, the witness was shown an evidence (i.e., a photograph) and instructed randomly by the ADDR system to be either truthful or deceptive about describing the image. The interrogator was guided by the system to ask the witness specific questions to determine whether the witness was being truthful or not. The participants thus took turns to speak during the conversation. They were motivated by payments to follow their roles. The witness received a bonus only if the interrogator believed him/her.

TABLE I
'Relevant questions' used in the ADDR framework

What was your image?
Could you give me some more details about the image?
If there were something to count in the image, what would it be and what would be the count?
Were there any other objects in the image?
What were the colors in the image?
Please tell me about the background in the image.
Where do you think the photograph was taken?
Were parts of the object in the image man-made?

The interrogation game comprised of two main components: (1) baseline questions and (2) relevant questions. The baseline questions were used to gauge a witness's behavior when speaking about topics unrelated to the image. During the relevant questioning, the interrogator was prompted by the ADDR system to ask the witness a set of questions related to the image, as shown in Table I. In this experiment, we only use the relevant questions.

The data captured using this framework and interrogation protocol include $N = 151$ dyads (Truth $= 75$, Lie $= 76$) of video data (1.2 million frames). Each interaction was approximately 5 minutes in duration and was recorded at 15 frames per second. In this analysis, we only use the video frames involved with the witnesses. Participants (40% female) included both Amazon Mechanical Turkers and University students. We manually transcribed all the videos to perform linguistic analysis. The total word count of all the transcripts is around $42,000$.

III. Research Question

Our research is focused on the following question: *Does looking into language patterns in light of facial expression contexts reveal any meaningful insight in understanding deceptive behavior?*

IV. Methods

This section is divided into three parts. First, we identify linguistic and facial expression cues/features *without* considering any context, towards distinguishing truth-tellers from deceivers. Second, we explain how our framework establishes context and the features associated with the process. Finally,

TABLE II
EXAMPLES OF LIWC CATEGORIES

Category	Examples
Perceptual Processes	Observing, heard feeling.
Seeing	View, Saw, Seen
Hear	Listen, Hearing
Cognitive Processes	Cause, know, ought
Causation	Because, effect, hence
Inclusive	And, with, include
Exclusive	But, without, exclude

we describe the statistical methods that are used to differentiate between deceivers and truth-tellers. The pipeline of the framework is summarized in Figure 1.

A. Analyzing Deceptive Behavior Without Context

Keeping the notion of context aside, we first identify and extract the verbal and non-verbal cues from the dataset, as described below.

1) Identification of Key Linguistic Features: We analyze the transcripts of only the witnesses, and refer to their responses as 'turns' or 'answers'. We use the Linguistic Inquiry and Word Count (LIWC) dictionary [22] to extract linguistic features from the answers. The dictionary consists of around $4,486$ words, divided across 64 dimensions or contextual categories. Many of these categories are strongly associated with psychometric properties, such as positive and negative emotions, cognitive and perceptual processes etc. LIWC also identifies language composition categories including prepositions, pronouns, articles, etc. Researchers have widely adopted LIWC for analyzing human communication behavior towards understanding personality styles [23], social media behavior [24] and deception [5], [25], [26]. This success is despite the fact that the LIWC dictionary only consists of $4,486$ words/word segments and that typical analysis only involves counting words in the text that appear in the dictionary. Examples of some LIWC category words are shown in Table II [22].

In each answer of the witnesses, we count the number of spoken words associated with each LIWC category. Then we take the average count per answer over the entire interrogation game video. We use these average counts as linguistic features to find key LIWC indicators for distinguishing deceptors from truth-tellers.

2) Identification of Key Facial Action Unit Features: To analyze the witness's facial expressions, we use OpenFace, an open source toolbox [34]. It gives the 2D and 3D coordinates of 68 landmarks on the face, gaze direction, head pose and facial Action Units (AU) based on the Facial Action Coding System (FACS) [12]. The performance of OpenFace has been benchmarked on manually coded public databases [34]. We remove all frames where face tracking is not successful. In addition, we only consider frames where tracking confidence is over 90%. In this study, we use the intensity scores of the 17 AU features, which lie within a range of $0-5$. The videos are recorded at 15 frames per second. We take the average AUs over all the answers in each video. These average AUs are used in further significance tests to identify key AU features for deception analysis.

B. Analyzing Deceptive Behavior With Facial Expression Based Context

We use our framework to combine facial expression contexts automatically with linguistic features. We extract the AUs and LIWC category counts in a similar way as Sections IV-A1 and IV-A2. In the sequel, we build on those objective measures.

1) Identify Video Regions Associated With High and Low Intensities of Facial Action Units: We take the average of the different AUs over all frames corresponding to each answer by the witnesses. Then, for each AU, we calculate the median of these average values over all the answers in the entire dataset. We define an answer to have a 'high intensity' in an AU if the average AU intensity in that answer is higher than the median value. Otherwise, the answer has a 'low intensity' of that AU. For example, in step 4 of Figure 1, we mark a couple of red regions in the video with '*' and '**'. During the answer marked by '**', the average of AU05 (Upper Lid Raiser) is greater than the median value of average AU05's computed over the entire dataset. In other words, the witness expressed a high intensity of AU05 in that answer. On the contrary, the witness expressed a low intensity of AU05 (<median) in the '*' marked answer. Our framework automatically labels all the answers on whether they have high or low intensities of different Action Units.

2) Extracting LIWC Features Based on Facial Action Unit Context: For each answer, we count the number of words corresponding to each LIWC category. Then, we consider the answers corresponding to 'high intensity' of each AU consecutively, and take the averages for different LIWC category counts in those answers. We repeat the same for the 'low intensity' answers. For example, in step 4 of Figure 1, we find all the answers associated with high (marked with **) and low (marked with *) intensities of AU05. This is the context for this AU. In step 5, we extract all the answers associated with these marked regions. Then, we count the number of 'Perceptual' words present in those answers, as an example of a LIWC feature. We then take the average of Perceptual word counts for the answers corresponding to the '*' and '**' segments, and use these values in the analyses that follow. We repeat the same process for all the AUs, and in each of those, all the LIWC features. Finally, significance tests are done on the extracted features to identify differences between truth-tellers and bluffers in both regions.

C. Statistical Methods

We use the Student's unpaired t-test to analyze the hypothesis–whether the extracted verbal and non-verbal features are different for the deceivers and truth-tellers. The t-test assumes that each of the distributions are normally distributed which may or may not be the case. As a result, we also use the Mann-Whitney-Wilcoxon test (MWW) [35], which does not make the normal distribution assumption. Cohen's d [36] is used to quantify the effect size, namely, the difference

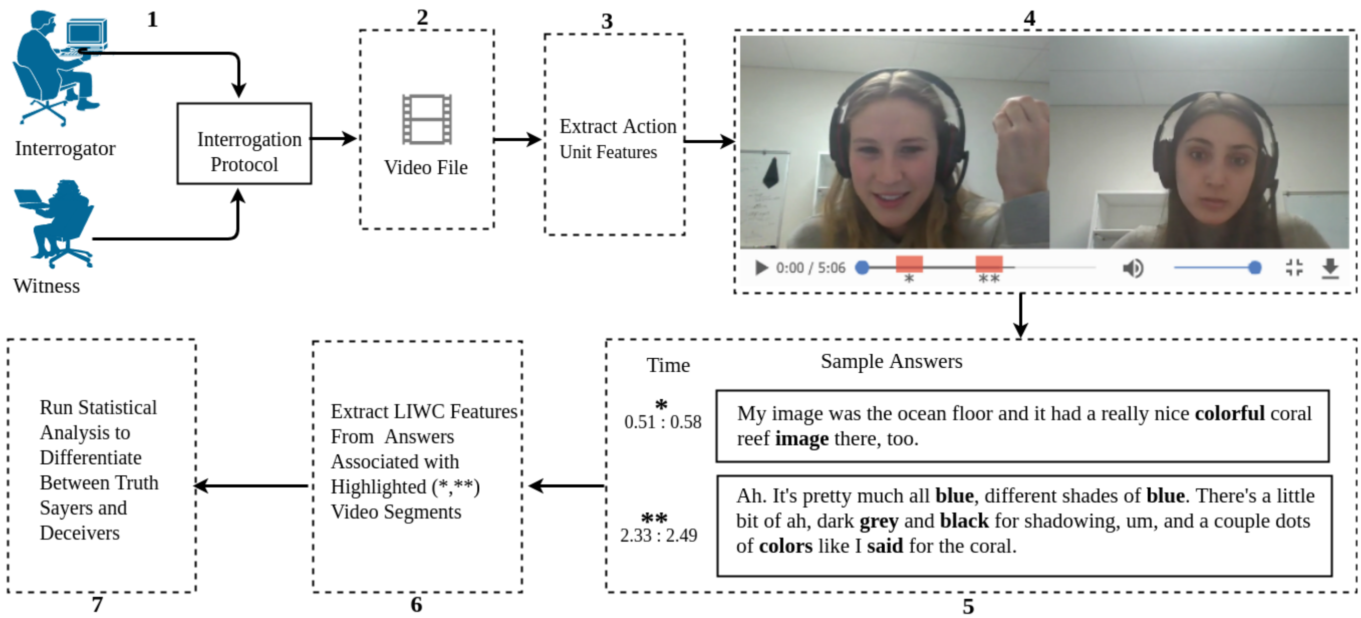

Fig. 1. Pipeline of our framework. 1) An interrogator questions a witness following an interrogation protocol, and tries to figure out whether the witness is being deceptive or not. 2) The video is recorded. 3) Facial Action Units are extracted from the video frames. 4) Video segments are identified where the witness displays high intensities of Action Units. For example, an answer associated with a high intensity of AU05 (upper lid raiser) is marked with a '**' in the figure, and the corresponding time-frames in red. Similarly, we marked a low intensity region with '*'. 5) The answers corresponding to the marked regions are tracked. (6) All LIWC categories used in the answers are identified. For example, all highlighted words in step 5 are associated with the LIWC category 'Perceptual'. 7) Extracted LIWC patterns are analyzed for understanding deceptive behavior.

between the truth-tellers' and bluffers' feature scores in units of estimated standard deviation. For each experiment, we test the same hypothesis for all verbal and non-verbal features. Bonferroni correction [37] is done to adjust the significance values, as detailed in the Results section.

V. Results

This section is organized into two main parts. In subsections V-A and V-B, we report the outcomes of the statistical analyses without and with contexts, respectively.

A. Exploratory Study Without Considering Context

We find that without any context, the LIWC features on their own can show statistical differences between truth-tellers and bluffers. However, facial expression features alone cannot result in such differences.

1) Results of Linguistic Feature Analysis: The results of the statistical analyses of LIWC features are listed in Table III. We only present the key LIWC categories that show statistically significant differences between truth-tellers and bluffers after Bonferroni correction ($p < 0.05$). Truth-tellers use more 'Perceptual', 'Seeing' and 'Cognitive' LIWC category words compared to deceivers. The Cohen's d values also indicate that the standardized differences between the pairs of means are notable ($d > 0.62$) and greater than the medium effect size (0.5) [36]. From Table II, we can see that 'Seeing' is a subcategory of the perceptual process [22]. In addition, we find that, on average, the deceivers use less number of words per answer than truthful people ($p < 0.05$, truth mean$= 300.34$ and bluff mean$= 265.59$).

TABLE III
Statistical Analysis Results of LIWC Categories Between Truth-Tellers and Bluffers

Category	Mean Truth	Mean Bluff	t-test	MWW	Cohen's d
Seeing	0.86	0.44	**0.00018**	**0.00006**	**-0.65**
Perceptual	1.11	0.57	**0.00019**	**0.00010**	**-0.65**
Cognitive	4.25	2.77	**0.00027**	**0.00079**	**-0.63**

2) Results of Facial Action Unit Analysis: We do not find any statistically significant differences in facial AU features. None of the video-wide average AU values can differentiate the deceivers from the truth-tellers. This is intuitive, as a summary metric (i.e., average) from all of the frames may not always capture the fine-grained dynamics of the AUs in the deceptive moments of interest.

B. Exploratory Study With Context

We find that analyzing linguistic differences in light of facial AUs reveal significant distinctions between truth-tellers and bluffers, occasionally with stronger signals than using linguistic features alone.

1) Median Values for Facial Action Unit Context: We explained in subsection IV-B1 how we calculate the median intensity for each AU. These median values are listed in Table IV. We use these values to label the answers to be of high or low intensities of different AUs. For example, if an

TABLE IV
MEDIAN AU VALUES FOR DIFFERENT AUS

Action Units	Median	Action Units	Median
1	0.21	14	0.54
2	0.38	15	0.08
4	0.17	17	0.44
5	0.04	20	0.07
6	0.38	23	0.09
7	0.64	25	0.82
9	0.08	26	0.55
10	0.60	45	0.28
12	0.54		

TABLE V
STATISTICAL ANALYSIS RESULTS OF LIWC FEATURES BASED ON FACIAL AU CONTEXTS; H=HIGH INTENSITY REGIONS, L=LOW INTENSITY REGIONS, S=SEEING, P=PERCEPTUAL, C=COGNITIVE CATEGORIES

AU Freq.	LIWC Cat.	Mean Truth	Mean Bluff	t-test	MWW	Cohen's d
AU05-H	S	0.97	0.49	**2.46E-5**	**8.76E-6**	**-0.75**
AU05-H	P	1.30	0.66	**5.99E-5**	**3.28E-5**	**-0.71**
AU05-H	C	4.94	3.01	**4.20E-5**	**9.08E-5**	**-0.72**
AU14-H	P	1.11	0.38	0.0001	6.93E-5	-0.75
AU07-L	S	0.91	0.40	6.49E-5	0.00015	-0.77
AU12-H	C	4.52	2.49	0.0001	0.00011	-0.74

answer has an average AU05 value greater than or equal to 0.04, then we label that answer as a high intensity instance of AU05.

2) Results of LIWC Pattern Analysis Based on Facial Action Unit Context: As mentioned in section IV-B2, for each AU, we look at both the high and low intensity answers. We calculate the average LIWC category counts in those regions and run statistical analyses to explore the differences between truth-tellers and deceivers. The top statistically significant results are listed in Table V ($p < 0.05$). The p-values for the t-test and MWW test are shown in the table. When we adjust these values by Bonferroni correction (17 action units $\times$ 64 LIWC categories $\times$ 2 intensity regions), the MWW test p-value for the 'Seeing' category during high intensity regions of AU05 appears significant. The t-tests and MWW tests for 'Seeing', 'Perceptual' and 'Cognitive' categories miss Bonferroni corrected $p < 0.05$ significance by a small margin in the high intensity AU05 regions. The effect size, Cohen's d, is also notable and close to the greater effect size ($d = 0.8$) [36]. In all cases, the deceivers produce less Seeing, Perceptual and Cognitive words, which is consistent with our findings in section V-A1.

We saw earlier in Table III that the LIWC categories Seeing, Perceptual and Cognitive can distinguish between truth-tellers and bluffers significantly. However, if we split the answers into two parts based on high and low intensity AU contexts, in some cases the effect becomes stronger in either intensity region and weaker in the other, and in some other cases they show similar effects in both intensity regions. In Table V, we see that the witnesses show stronger differences in their use of Seeing, Perceptual and Cognitive words in the regions of high intensity of AU05 than those seen in Table III. Whereas, their counterpart low intensity regions of AU05 show no significant difference at all, as reported in Table VI. Further, it can be seen from Table VI that the counterpart regions of AU 14, 07 and 12 show significant differences between truth-tellers and bluffers (without Bonferroni correction). Figure 2 illustrates the comparison of Seeing category usage between high and low intensity AU05 regions.

TABLE VI
STATISTICAL ANALYSIS RESULTS OF LIWC FEATURES BASED ON FACIAL AU CONTEXTS; H=HIGH INTENSITY REGIONS, L=LOW INTENSITY REGIONS (COUNTERPART REGIONS TO TABLE V)

Partition	Category	Mean Truth	Mean Bluff	t-test	MWW	Cohen d
AU05-L	Seeing	0.75	0.45	0.09	0.48	-0.28
AU05-L	Perceptual	0.89	0.52	0.05	0.49	-0.33
AU05-L	Cognitive	3.27	2.55	0.12	0.10	-0.26
AU14-L	Perceptual	0.88	0.62	0.03	0.03	-0.41
AU07-H	Seeing	0.71	0.36	0.02	0.01	-0.48
AU12-L	Cognitive	3.76	2.47	0.003	0.01	-0.54

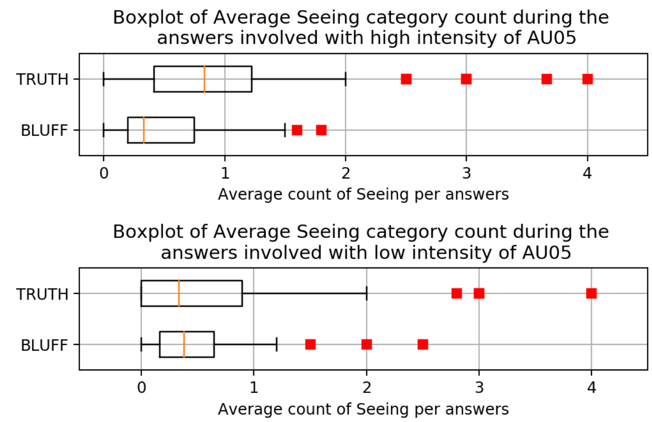

Fig. 2. Comparison of Seeing category usage between high and low intensity AU05 regions

We also look at the number of words per answer by deceivers and truth-tellers. The results are presented in Table VII. Without taking into account the AU contexts, we can see that the deceivers produce significantly shorter answers (first row of Table VII; t-test $p = 0.001$, MWW $p = 0.002$ and $d = -0.55$). Table VII also includes the average word count per answer during the high and low intensity regions for the key AU features reported in Table V. In all cases the truth-tellers produce significantly more words than bluffers, with the exception of low intensity regions of AU05 ($p > 0.5$). However, only the high intensity regions of AU05 showed a stronger p-value than that without context. Figure 3 further illustrates the point.

VI. DISCUSSION

In Section V-A1, we found that the LIWC categories 'Seeing', 'Perceptual' and 'Cognitive' can differentiate the truth-tellers from bluffers significantly without any contextual cues from AUs. These results corroborate findings in previous literature, where bluffers were shown to use fewer words associ-

TABLE VII
COMPARISON BETWEEN TRUTH-TELLERS AND DECEIVERS ON AVERAGE NUMBER OF WORDS PER ANSWER; H=HIGH INTENSITY REGIONS, L=LOW INTENSITY REGIONS

Partition	Mean Truth	Mean Bluff	t-test	MWW	Cohen d
Without context	27.95	19.77	**0.001**	**0.002**	**-0.55**
AU05_H	33.20	22.17	**0.0003**	**0.0004**	**-0.63**
AU05_L	20.67	17.61	**0.25**	**0.25**	**-0.19**
AU07-H	25.54	16.54	0.001	0.002	-0.65
AU07-L	27.87	18.27	0.003	0.01	-0.56
AU12-H	28.49	18.38	0.001	0.001	-0.60
AU12-L	24.75	18.10	0.01	0.03	-0.45
AU14-H	27.62	17.10	0.001	0.001	-0.63
AU14-L	23.63	19.60	0.12	0.11	-0.30

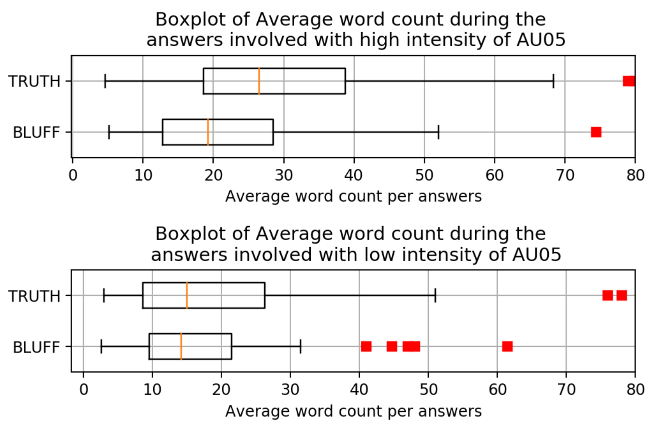

Fig. 3. Comparison of average number of words per answer between high and low intensity AU05 regions

ated with the cognitive [6] and perceptual processes [38] compared to truthful participants. The witnesses used perceptual and cognitive words mostly to explain the given image during interrogation. Fabricating a false story is a difficult task and that can result in producing fewer words by the bluffers [39]. From our results, we also see that bluffers produced fewer words than truth-tellers (Section V-B2). Specifically, bluffers tend to avoid perceptual and cognitive words, which are keys to describing an image.

Therefore, LIWC categories alone can distinguish truth-tellers and bluffers well. Analyzing the LIWC categories on top of facial AU-based contexts give further insights into such distinguishing, towards answering our research question. We find that when we split the answers based on high or low AU intensities, in some cases the distinguishing significance becomes stronger in either intensity region and weaker in the other, while in some other cases both intensity regions give similar significance in distinguishing truth-tellers from bluffers. Interestingly, the average facial AUs alone did not gave any meaningful difference, as described in Section V-A2. This is also reasonable if one considers the micro-expression theory by Paul Ekman [10]. Micro-expressions are naturally very short in duration. Therefore, the information contained in those short occurrences of AUs might get lost once averaged over the entire video. Our framework automatically gives attention to the important micro-expression AU moments to set the contexts for linguistic analysis.

We found meaningful insights on linguistic patterns based on AU05 context (upper lid raiser/eye open). The differences in the usage of 'Seeing', 'Perceptual' and 'Cognitive' LIWC categories are larger in significance when we isolate our analyses to the high intensity regions of AU05, compared to their context-agnostic counterparts. On the other hand, the low intensity regions do not show any differences. This same pattern is also observed in average word count per answer. These effects are only visible when attention is given to the important micro-expression AU segments by our framework.

Making eye contact or pretending to remember something can be a difficult task for deceivers in fabricating their story. For example, Vrij *et al.* [27] found that it is easier for interrogators to distinguish between truth-tellers and deceivers when eye contact is made. As a result, looking into the eyes may hold a key to understanding the emotions a person is going through.

We used the largest interrogation dataset that is publicly available for our study. The limitation of this dataset is that it consists of unsanctioned lies and that the stake is low. The witnesses were not able to choose whether to tell the truth or not. This decision was made randomly by the ADDR framework. The participants did not get enough preparation time to pre-script their answers. Therefore, deceptive behavior in a sanctioned situation, where people have the choice of telling the truth or lying, could potentially be different. There is also no punishment for the participants if the interrogator caught them lying. In real life scenarios, like border customs or criminal interrogations, there is usually a price to pay for lying. Lying in those scenarios will be more stressful. Deceptive behavior patterns might be different in such higher-stake real life scenarios. In the future, we wish to apply our framework to compare sanctioned vs. unsanctioned lies, as well as well high-stake vs. low stake lies, during deceptive communication.

VII. CONCLUSION

In this paper, we presented a framework that automatically establishes facial AU based contexts to enable subsequent linguistic analysis. Our analyses showed that these contexts are important to find new insights on deceptive behavior. During a highly intense display of AU05 (upper lid raised/eyes wide open), fewer words from the 'Seeing', 'Perceptual' and 'Cognitive' LIWC categories were used by the deceivers. In addition, deceivers tend to provide shorter answers in those AU regions compared to truth-tellers. On the contrary, these differences are absent when we look into the low intensity regions of AU05. The linguistic usage patterns were markedly different based on the contexts of high vs low intensities of facial expressions. Our framework can find these new insights automatically, which helps in understanding deception in multimodal communication.

REFERENCES

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

[2] L. Caso, F. Maricchiolo, M. Bonaiuto, A. Vrij, and S. Mann, "The impact of deception and suspicion on different hand movements," *Journal of Nonverbal Behavior*, vol. 30, no. 1, pp. 1–19, 2006.

[3] A. K. Webb, C. R. Honts, J. C. Kircher, P. Bernhardt, and A. E. Cook, "Effectiveness of pupil diameter in a probable-lie comparison question test for deception," *Legal and Criminological Psychology*, vol. 14, no. 2, pp. 279–292, 2009.

[4] S. Lu, G. Tsechpenakis, D. N. Metaxas, M. L. Jensen, and J. Kruse, "Blob analysis of the head and hands: A method for deception detection," in *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*. IEEE, 2005, pp. 20c–20c.

[5] J. T. Hancock, L. E. Curry, S. Goorha, and M. Woodworth, "On lying and being lied to: A linguistic analysis of deception in computer-mediated communication," *Discourse Processes*, vol. 45, no. 1, pp. 1–23, 2007.

[6] M. L. Newman, J. W. Pennebaker, D. S. Berry, and J. M. Richards, "Lying words: Predicting deception from linguistic styles," *Personality and Social Psychology Bulletin*, vol. 29, no. 5, pp. 665–675, 2003.

[7] A. Vrij, S. Leal, P. A. Granhag, S. Mann, R. P. Fisher, J. Hillman, and K. Sperry, "Outsmarting the liars: The benefit of asking unanticipated questions," *Law and Human Behavior*, vol. 33, no. 2, pp. 159–166, 2009.

[8] 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.

[9] 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 *Proceedings of the 4th International Conference on Affective Computing and Intelligent Interaction - Volume Part I*, ser. ACII'11. Berlin, Heidelberg: Springer-Verlag, 2011, pp. 135–144. [Online]. Available: http://dl.acm.org/citation.cfm?id=2062780.2062799

[10] P. Ekman and W. V. Friesen, "The repertoire of nonverbal behavior: Categories, origins, usage, and coding," *Semiotica*, vol. 1, no. 1, pp. 49–98, 1969.

[11] G. Warren, E. Schertler, and P. Bull, "Detecting deception from emotional and unemotional cues," *Journal of Nonverbal Behavior*, vol. 33, no. 1, pp. 59–69, 2009.

[12] P. Ekman and W. V. Friesen, "Facial action coding system," 1977.

[13] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The Extended Cohn-Kanade dataset (CK+): A complete dataset for action unit and emotion-specified expression," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*. IEEE, 2010, pp. 94–101.

[14] P. Ekman, W. V. Freisen, and S. Ancoli, "Facial signs of emotional experience," *Journal of Personality and Social Psychology*, vol. 39, no. 6, p. 1125, 1980.

[15] P. Ekman and W. V. Friesen, "Felt, false, and miserable smiles," *Journal of Nonverbal Behavior*, vol. 6, no. 4, pp. 238–252, 1982.

[16] R. El Kaliouby and P. Robinson, "Real-time inference of complex mental states from facial expressions and head gestures," in *Real-time Vision for Human-Computer Interaction*. Springer, 2005, pp. 181–200.

[17] ——, "Mind reading machines: Automated inference of cognitive mental states from video," in *IEEE International Conference on Systems, Man and Cybernetics*, vol. 1. IEEE, 2004, pp. 682–688.

[18] T. Baltrušaitis, D. McDuff, N. Banda, M. Mahmoud, R. El Kaliouby, P. Robinson, and R. Picard, "Real-time inference of mental states from facial expressions and upper body gestures," in *IEEE International Conference on Automatic Face & Gesture Recognition (FG)*. IEEE, 2011, pp. 909–914.

[19] K. Fukuda, "Eye blinks: New indices for the detection of deception," *International Journal of Psychophysiology*, vol. 40, no. 3, pp. 239–245, 2001.

[20] R. I. Thackray and M. T. Orne, "A comparison of physiological indices in detection of deception," *Psychophysiology*, vol. 4, no. 3, pp. 329–339, 1968.

[21] C. R. Bartol and A. M. Bartol, *Psychology and law: Theory, research, and application*. Thomson/Wadsworth, 2004.

[22] J. W. Pennebaker, M. E. Francis, and R. J. Booth, "Linguistic inquiry and word count: LIWC," *Mahway: Lawrence Erlbaum Associates*, vol. 71, 2001.

[23] J. W. Pennebaker and L. A. King, "Linguistic styles: Language use as an individual difference," *Journal of Personality and Social Psychology*, vol. 77, no. 6, p. 1296, 1999.

[24] E. Gilbert and K. Karahalios, "Predicting tie strength with social media," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2009, pp. 211–220.

[25] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, "Finding deceptive opinion spam by any stretch of the imagination," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies - Volume 1*. Association for Computational Linguistics, 2011, pp. 309–319.

[26] Y. R. Tausczik and J. W. Pennebaker, "The psychological meaning of words: LIWC and computerized text analysis methods," *Journal of Language and Social Psychology*, vol. 29, no. 1, pp. 24–54, 2010.

[27] A. Vrij, S. Mann, S. Leal, and R. Fisher, "'Look into my eyes': Can an instruction to maintain eye contact facilitate lie detection?" *Psychology, Crime & Law*, vol. 16, no. 4, pp. 327–348, 2010.

[28] M. Abouelenien, V. Pérez-Rosas, R. Mihalcea, and M. Burzo, "Deception detection using a multimodal approach," in *Proceedings of the 16th International Conference on Multimodal Interaction*. ACM, 2014, pp. 58–65.

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

[30] V. Pérez-Rosas, M. Abouelenien, R. Mihalcea, Y. Xiao, C. Linton, and M. Burzo, "Verbal and nonverbal clues for real-life deception detection," in *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, 2015, pp. 2336–2346.

[31] 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.

[32] D. McDuff, P. Thomas, M. Czerwinski, and N. Craswell, "Multimodal analysis of vocal collaborative search: A public corpus and results," in *Proceedings of the 19th ACM International Conference on Multimodal Interaction*. ACM, 2017, pp. 456–463.

[33] 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.

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

[35] J. L. Devore, *Probability and Statistics for Engineering and the Sciences*. Cengage Learning, 2015.

[36] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*. Routledge, 1988.

[37] F. Bretz, P. Westfall, and T. Hothorn, *Multiple comparisons using R*. Chapman and Hall/CRC, 2016.

[38] G. D. Bond and A. Y. Lee, "Language of lies in prison: Linguistic classification of prisoners' truthful and deceptive natural language," *Applied Cognitive Psychology*, vol. 19, no. 3, pp. 313–329, 2005.

[39] C. L. Toma and J. T. Hancock, "Reading between the lines: Linguistic cues to deception in online dating profiles," in *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*. ACM, 2010, pp. 5–8.