Gender Differences in Multimodal Contact-Free Deception Detection

Mohamed Abouelenien
University of Michigan, Dearborn

Mihai Burzo
University of Michigan, Flint

Verónica Pérez-Rosas
University of Michigan, Ann Arbor

Rada Mihalcea
University of Michigan, Ann Arbor

Haitian Sun
University of Michigan, Ann Arbor

Bohan Zhao
University of Michigan, Ann Arbor

***Abstract*—In this paper, we explore the hypothesis that multimodal features as well as demographic information can play an important role in increasing the performance of automatic lie detection. We introduce a large, multimodal deception detection dataset balanced across genders, and we analyze the patterns associated with the thermal, linguistic, and visual responses of liars and truth-tellers. We show that our multimodal noncontact deception detection approach can lead to a performance in the range of 60%–80%, with different modalities, different genders, and different domain settings playing a role in the accuracy of the system.**

Automatic deception detection has been recently receiving an increasing amount of attention, due to the increase in security threats, and also motivated by the growth of online interactions that often include deceptive communication.[1] Most of the computational work to date, however, has focused on building general models that do not account for demographic information such as gender or age. Males and females vary in representing themselves and usually act differently, especially when it comes to deception.[2] Research work on deception detection has identified deceptive behaviors that can be attributed to gender differences.

For instance, females tend to lie about money and eating behavior whereas males tend to lie about sports and their jobs.[3] Taking this

Digital Object Identifier 10.1109/MMUL.2018.2883128
Date of publication 23 November 2018; date of current version 27 August 2019.

information into account, there are cases where the gender of the deceiver can be easily identified. "I am almost ready", "It was on sale," "No, You Do not Look Big in That," and "This will be my last drink." Based on a research study, it can be easily predicted that the first two statements are often made by females and the last two are usually told by males (http://datafication.com.au/). In most of the cases, judgments are not as easy as several other aspects come into play during deceptive behavior. Certain behaviors and patterns can be associated with specific gender, culture, or age.

Previous work in the field of psychology has explored the role played by gender in deception detection. For example, it was found that males have less ability to detect deceit compared to females.[4] None of this work however has considered the possibility of having gender-based variations in the automated deceit detection process.

In our previous work,[5] we introduced a multimodal approach for deception detection regardless of gender using 30 subjects, where we integrated linguistic and thermal features, along with physiological measurements extracted from contact-based sensors. In our recent research,[6] we used a contact-based approach to construct separate machine learning models for males and females in order to detect deception, where we also used the linguistic, thermal, and contact-based physiological modalities. This paper extends our previous work and presents an automated multimodal contact-free deception detection approach that accounts for gender differences, along with a novel in-depth analysis of the specific visual, linguistic, and thermal behaviors associated with each gender when they act deceptively. We no longer utilize the contact-based physiological sensors and instead use visual cameras to add the visual modality to the thermal and linguistic data streams.

Specifically, this paper makes three contributions, in addition to a large deception detection dataset that includes 520 instances from 104 participants, with approximately equal number of males and females. First, we extract and integrate features from three different data sources including verbal, visual, and thermal modalities to provide a noncontact approach to detect deceit. Second, we present a novel in-depth analysis of the behaviors and trends that are specific to each gender when it comes to deception for each of the linguistic, visual, and thermal modalities. Finally, we analyze the potential of using demographic information in improving the performance of our classification models to detect deception, and we compare the performance of our multimodal system to that of human annotators.

RELATED WORK

Verbal Deception Detection

Linguistic analysis was the focus of significant recent research work, owing to its noninvasiveness and its promising results in revealing clues of deception in a variety of domains where computer mediated communication occurs, including chats, forums, online dating, etc.[3,7]

Nonverbal Deception Detection

Gesture and facial expressions were also found to provide useful clues for deception detection. Spontaneous facial expressions and hand gestures were of special interest due to their usage to express people's emotions on daily basis.[8] For instance, participants used less gestures when they told a story in a deceptive manner compared to telling a similar story in a truthful manner.[9]

Several efforts were additionally exerted in the direction of noninvasive approaches of detecting deception using thermal imaging, leveraging changes in blood flow in the face.[10]

Gender-Based Deception Detection

Gender differences in deception have been studied in different fields such as psychology, economics, and linguistics among others. Economic studies showed that male participants tend to be more deceptive to achieve monetary benefits,[11] or that men are more likely to deceive others if it would result in additional benefit.

Males lied more often in online dating,[12] especially regarding their appearance and personality compared to females.

DATASET COLLECTION

A dataset of 520 deceptive and truthful responses was collected from 104 subjects using a setup consisting of a thermal camera, two visual cameras, and a microphone.

A FLIR thermal camera that has a resolution of 640×512 (FLIR SC6700) was used to capture thermal video recordings for the subjects. We also used two visual cameras, a Mightex scientific camera focusing on the facial area and a Logitech camera focusing on the upper body. Subjects' verbal responses were recorded separately using a noise canceling microphone.

Scenarios

We designed three scenarios for the participants. Two of the scenarios included statements that were made by the participants, one time in a truthful manner and a second time deceptively. In the third scenario, an interviewer questioned the participants and they chose whether to respond truthfully or deceptively. The scenarios are as follows:

Abortion

The subjects made a statement about their truthful opinion on abortion in their own words and the reasons behind taking a specific stance on this matter. This was followed by a second statement, where the subjects reversed their opinion and responded deceptively.

Best Friend

The participants started this scenario by making a truthful statement about their best friend, which might include shared memories, events, or incidents. This was followed by a deceptive statement made by the participants, where they describe someone they dislike positively as if this person was a best friend.

Mock Crime

The interviewer hid a $20 bill in a box inside a lab. The participants were then given the choice to steal the bill while the interviewer leaves the lab. After his return, he interviewed the participant and he or she chose whether to lie in their responses to the questions. The questionnaire used during the interview is shown below:

1. Are the lights on in this room?
2. Regarding that missing bill, do you intend to answer each question truthfully about that?
3. Prior to 2016, did you ever lie to someone who trusted you?
4. Did you take that bill?
5. Did you ever lie to keep out of trouble?
6. Did you take the bill from the private area of the lab?
7. Prior to this year, did you ever lie for personal gain?
8. What was inside the white envelope?
9. Please describe step by step, in as much detail as you can, what you did while you were in the room and I was outside.
10. Do you know where that missing bill is right now?

Table 1. Distribution of the choice of the 104 subjects on whether to lie or be truthful in the "Mock Crime" scenario as well as the interviewer correct prediction rate.

	Male	Female
Truthful	30 (58.8%)	27 (51%)
Deceptive	21 (41.2%)	26 (49%)
Overall No. of Responses	51	53
Interviewer prediction	45.1%	73.6%

After the interview was over, the interviewer predicted whether the participant stole the bill. If the interviewer was not correct, the participant would receive a higher incentive.

While an approximately similar number of females opted to respond truthfully and deceptively to the questions in the "Mock Crime" interview, a smaller percentage of males decided to respond truthfully as shown in Table 1. An interesting observation can also be noted that the interviewer was clearly capable of predicting lies in females compared to males with a relative improvement of 63.2%.

MULTIMODAL NONCONTACT FEATURE EXTRACTION

Verbal Cues of Deception

We start by transcribing the participants' statements via crowd-sourcing with Amazon Mechanical Turk (AMT). The workers transcribed the subjects' statements using the audio

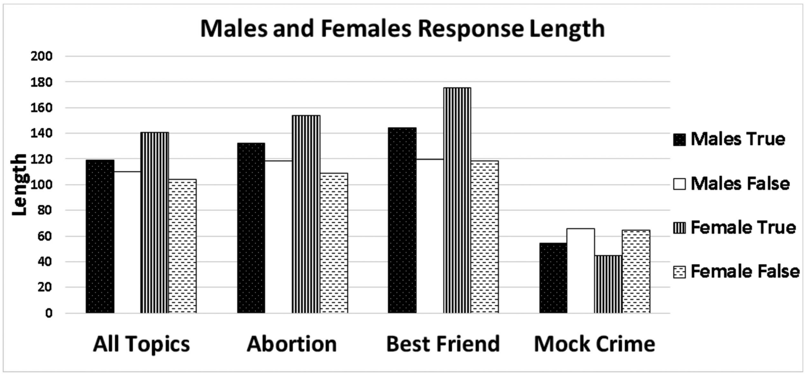


Figure 1. Length of truthful and deceptive responses from males and females for individual and combined topics.

recordings only. Transcripts include word filler and repetitions, as well as long pauses.

Figure 1 illustrates the length of the responses averaged for the males and females groups for individual and combined topics. It can be seen that both females and males have longer responses when they speak truthfully in "Abortion" and "Best Friend". The opposite holds for "Mock Crime," which could be due to an increase in the defensive words when the subjects acted deceptively during the interview. Interestingly, over all the topics, the average gap between the length of the truthful and deceptive responses for females is significantly larger than that for males, which indicates that in general verbal differences exist between females and males as they act deceptively.

We extracted features that were previously found to indicate deceptive behavior.[13]

Uni-grams: We used unigrams to represent the frequency of occurrence of unique words in the statements made by the subjects, which we derived from the bag-of-words representation.

LIWC features: LIWC includes 80 word classes to represent thoughts, personality, emotions, and motivations. We extracted our features using LIWC to classify the words in the transcripts of the subjects into their proper classes. Examples of such classes include "Other," which contains words, such as she, they, and he, and "I," which contains self-references, such as myself.

For additional insight, Figure 2 presents the top ranked semantic LIWC classes associated with deceptive and truthful statements, using the semantic word class scoring from.[14] The bars are calculated by subtracting the average frequency of words of the deceptive instances belonging to each class from the corresponding truthful instances for each gender. Hence, a positive result indicates an association between a LIWC class and truthfulness, and a negative result indicates an association between a LIWC class and

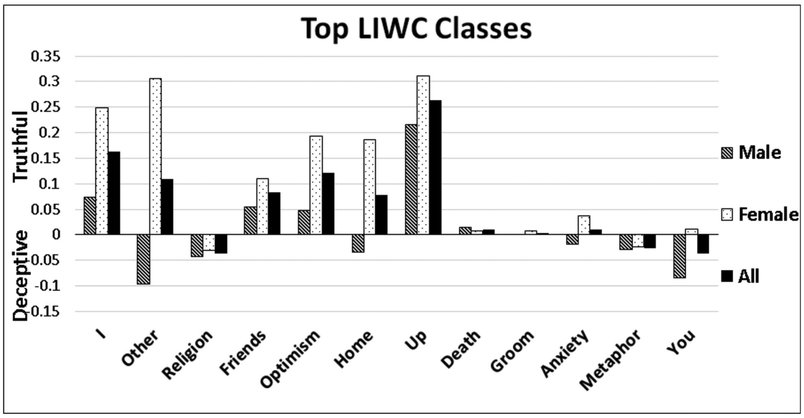


Figure 2. Gender differences in deception, as reflected in LIWC classes with top scores.

deception. Interestingly, the figure indicates that unlike females, males try to distant themselves from an action using words related to “Other” and “You” when they lie. Additionally, the “Home” and “Anxiety” classes have stronger association with lying for males. Moreover, the majority of the classes associated with truthfulness for both genders exhibit higher values for females.

Syntactic complexity and readability features: We extracted syntactic complexity and readability features,[15] as well as readability metrics. The features contained 14 indexes that represented the degree of syntactic complexity of the responses of the subjects. Examples of these features include coordinate phrases per unit and clause, clauses and dependent clauses per sentence and t-unit (the shortest grammatically allowable sentences into which text can be split), mean length of sentences, clauses, and t-units per sentence.

Shallow and deep syntax: The shallow and deep syntax features were extracted from part-of-speech (POS) tags and context free grammars trees (CFG) using the Stanford parser. We extracted POS features, which we represented as frequency values of all the tags in the transcripts. We also extracted production rules as frequency values using CFG,[7] which consisted of the lexicalized rules integrated with the grandparent node.

Length features: In order to take the temporal aspect of the responses into consideration, we divided the responses of the subjects into five intervals of similar length represented in terms of sentences and computed the number of words spoken in each interval. Accordingly, five features were extracted to indicate how the length of the responses of the subjects varied over time in order to explore whether certain patterns are indicative of deception.

Visual Gestures

Visual behavior has been found to be highly correlated to deceptive behavior. In order to incorporate information about visual behaviors that might be correlated to deception, we annotated subjects’ facial displays and hand movements using the MUMIN coding scheme.[16] Our choice of using MUMIN is motivated by the need of gesture annotations that accurately capture multimodal communication behaviors in our videos. MUMIN is an annotation scheme created for the annotation of multimodal communication of video clips in interview settings. It contains a coding scheme for gestures and facial displays in personal communication that are related to multimodal expressions.

We used nine categories from MUMIN, including eyebrow movements; eye movements; general facial expression; gaze direction; hand movement; hand trajectory; head movements; lips movements; and mouth openness.

To obtain accurate annotations, we label video segments of 20 seconds length; on average each video clip was split into four segments. The annotation is conducted using AMT, a crowdsourcing platform that has been successfully used in the past to obtain human annotations for different visual and text categorization tasks. AMT makes available on-demand workers who conduct a variety of annotation tasks. Each task is setup in the AMT web interface and made available for qualified workers. In our case, the worker qualifications included 1) the worker is located in the US, 2) the worker has an acceptance rate of at least 95% in previous submissions.

The gesture annotation task consists of labeling 40 different gestures in the facial and hand movement categories as described in the MUMIN scheme. To measure the reliability of the annotation task, we requested that three independent workers annotate each hit. Also, to ensure the quality of the annotation, i.e., avoid cases where annotator just select random labels for gesture annotation, we used the following strategies 1) randomly inserting a control video for which we knew the correct gesture label; 2) rejecting contributions where the elapsed time between accepting the task and submitting the response was less than the total duration of the videos the annotator was supposed to watch. Once the annotations were delivered, we assigned the final labels at the video level as well as at the segment level using majority voting over the labels given by the three annotators.

Visual Features From the gesture annotations at video level, we derived 40 binary features that

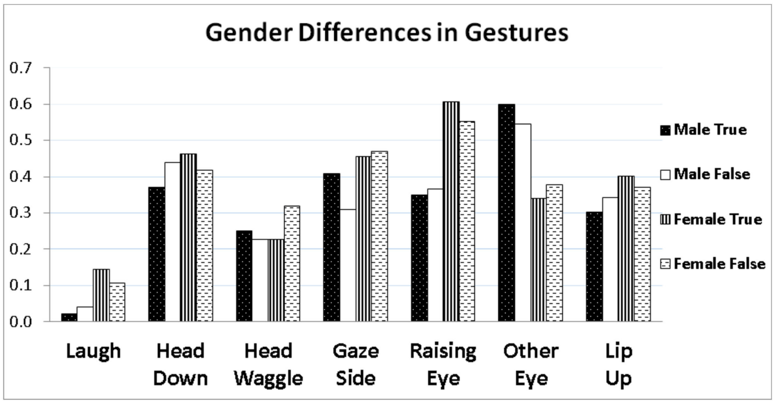


Figure 3. Differences in usage of certain gestures between males and females as they respond truthfully and deceptively.

represented the presence or absence of a given gesture while the subject was speaking truthfully or deceptively.

To identify differences in gesture behavior among male and female participants, we compared the percentage of each behavior as observed in each class. Figure 3 shows the percentages of all the visual gestures for which noticeable differences were identified between males and females for their deceptive and truthful responses. It can be noted that truthful females and deceptive males tend to laugh, have their heads directed downwards, raise their eyebrows, and have their lips shaped upwards during their responses. On the other hand, deceptive females and truthful males tend to waggle their heads and have a side gaze.

Moreover, temporal features were extracted from the visual gestures by calculating the number of times a gesture label changed in the video segments of each response. We extracted these features for each of the nine gesture categories in the MUMIN scheme.

The final list of visual features consisted of 40 binary features of hand and facial displays for each response and a set of nine features that indicated the dynamics of the gestures in the subjects' responses.

Thermal Responses

Segmenting and Tracking Regions of Interest (ROI) In order to analyze whether thermal differences occur between males and females as they act deceptively, first, we specified our ROI manually in the first frame of each response. The ROIs included the whole face, the forehead, the periorbital area, the cheeks with the nose, and the nose only. Second, interesting points were located in each ROI using the Shi-Tomasi corner detection algorithm. These points were detected in regions where there are sharper changes in the temperatures, which can potentially show whether an increase in the blood flow occurs as the subjects behaved deceptively. Third, the points were tracked throughout the video responses using a fast Kanade-Lucas-Tomasi tracking algorithm.

Once the tracking process was performed, interesting points were mapped from one frame to the next based on similarity by globally estimating their transition using geometric transformation. In order to ensure an accurate tracking process, we set the maximum distance between an interesting point and its location in the next frame to five and we set a threshold of 95% of matched points between successive frames.

Thermal Feature Extraction The locations of the bounding boxes containing the ROI of each frame were cropped from the raw thermal video, and a thermal map was created for each response to define the heat distribution in each ROI. In particular, statistical features were extracted from the responses on the video-level including the average, maximum, standard deviation, minimum, average of the frame-level standard deviation, and the mean of the 10% maximum temperatures in each ROI. A thermal

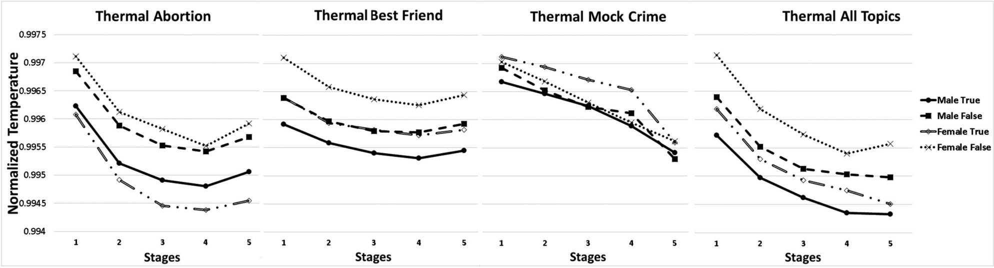


Figure 4. Normalized mean face temperatures of males and females divided into five stages throughout their truthful and deceptive responses for individual topics and all topics combined.

correction process was performed to account for the normal inter-personal temperature variations by dividing the features from the responses by the corresponding resting baseline features (from an additional resting recording).

Temporal features were also extracted by dividing each thermal video into five equal intervals, followed by computing statistical features from each of the intervals. This was performed in order to find gender-based thermal variations as the subjects responded deceptively. Figure 4 shows the "mean" feature averaged for the males and females groups for each of the five intervals. For "Abortion" and "Best Friend" topics, it can be clearly seen that both deceptive males and females exhibit an increase in their facial temperatures compared to truthful males and females. This trend is also reflected in the "All Topics" curves. However, it can be seen that in "Abortion," the deviation between the deceptive and truthful female curves is significantly larger than the male curves, which could be related to the nature of the topic that is more related to females.

Interestingly, this trend is different for the "Mock Crime" scenario. While deceptive males still exhibit an increase in temperature when they lie, deceptive females incur decreased temperatures compared to truthful females. This indicates that the perception of questions for the "Mock Crime" scenario was different between deceptive males and females. This could also be an indication that males took the "Mock Crime" scenario seriously, unlike females, which is reflected in the interviewer prediction results. Also, it agrees with the research concluding that males are better in lying when it comes to monetary benefits.[11]

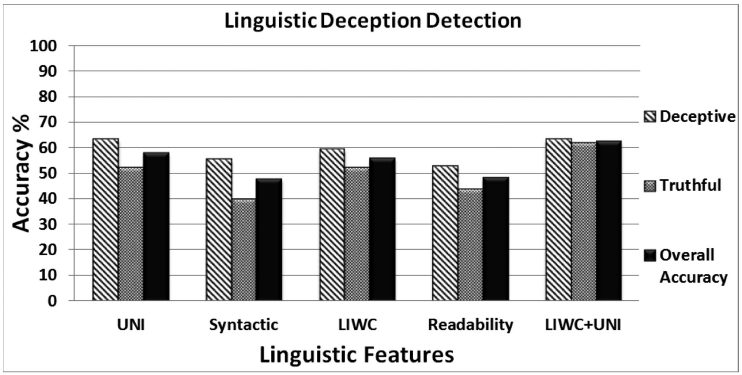
LEARNING DECEPTIVE BEHAVIOR

Following the feature extraction process, the classification process was conducted using features from individual and integrated modalities. We used a leave-one-subject-out cross validation scheme, where at each fold all instances of one subject were used for testing and all other instances from the other subjects were used for training. We chose to use a decision tree classifier, based on the superiority of this classifier over other classifiers for the task of deception detection, as recommended in previous work,[5,17,18] and as it provides better model interpretability by visualizing the nodes and levels of the tree. In addition, we used an SVM classifier with rbf kernel for comparison using all modalities combined. We report the overall average accuracy and recall of the deceptive and truthful classes for males and females separately and combined as well as for individual and combined topics.

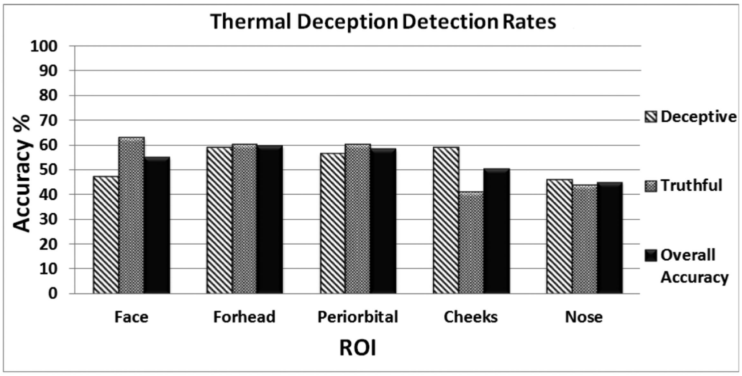
Development Data Results

In order to select the linguistic and thermal features that are most useful for deception detection, we performed evaluations and feature selection on a development dataset, which consists of recordings from 30 different subjects.[5] For our new experiments, we used this development dataset to specify which ROI is the most capable of indicating deceit using the thermal features as well as determine the best set of linguistic features that can differentiate between truthfulness and deception.

Figure 5(a) shows the overall accuracy and the recall using different sets of linguistic features. We also experimented different combinations of these sets and only showed the best possible combination. The figure indicates a



(a)



(b)

Figure 5. Overall accuracy and recall of the truthful and deceptive classes for (a) five sets of linguistic features and (b) the five thermal ROIs using the development dataset.

superior performance using the combination of the unigrams and LIWC features compared to using all other individual linguistic sets. Accordingly, we use this combined linguistic set for our new experiments.

Figure 5(b) shows the recall and overall accuracy of the different ROIs using the thermal features. The figure indicates that the forehead and periorbital regions provide the best capability of discriminating between deception and truthfulness. The forehead, in particular, attains the highest recall and overall accuracy, and hence is used in our new experiments.

EXPERIMENTAL RESULTS

The experiments reported in this section are conducted on the newly collected dataset of 520 instances. The dataset is gender-balanced and has a distribution of 255 deceptive and 265 truthful instances, with 255 responses from males and 265 from females.

Table 2 shows the average accuracy using all the visual features, as well as the linguistic (LIWC+Unigrams) and thermal (Forehead) features, which we select based on the development data results. We show the performance for individual and combined topics, as well as individual and combined genders using the decision tree classifier. For comparison, we also show the average accuracy using an SVM classifier with rbf kernel for "All Modalities" in the last column of the table. The majority baseline performance is listed in the first column for every set of experiments.

The table points to several interesting observations. First, it shows that overall the deceptive and truthful statements are easier to predict for females compared to males in individual and combined topics. This can be seen in 24 out of 32 cases. The exceptions are mostly found in the "Best Friend" scenario. Second, the integration of features from multiple modalities exhibits an improved performance in most cases compared to the usage of individual modalities except for the "Mock Crime" scenario, where the linguistic features performance stands out, and hence

Table 2. Overall accuracy for individual and combined features, genders, and topics using decision tree classifier, except for the last column, which uses SVM with rbf kernel.

	Baseline	Ling.	Thermal	Visual	Ling. + Thermal	Ling. + Visual	Thermal + Visual	All Modalities	(SVM) All Modalities
Abortion									
{Both}	50.0	59.1	53.8	54.8	**63.5**	58.2	56.7	63.0	54.8
{Male}	50.0	56.9	54.9	54.9	54.9	52.0	**59.8**	52.0	52.9
{Fem.}	50.0	63.2	56.6	61.3	57.5	61.3	**73.6**	57.5	62.3
Best Friend									
{Both}	50.0	54.8	50.5	54.3	**61.1**	53.8	56.7	56.7	59.6
{Male}	50.0	59.8	**69.6**	54.9	55.9	60.8	64.7	56.9	51.0
{Fem.}	50.0	46.2	55.7	55.7	47.2	48.1	54.7	47.2	**66.0**
Mock Crime									
{Both}	54.8	45.2	**59.6**	43.3	58.7	48.1	54.8	55.8	54.8
{Male}	58.8	**64.7**	51.0	62.7	56.9	62.7	54.9	62.7	47.1
{Fem.}	50.9	**81.1**	52.8	56.6	69.8	71.7	41.5	67.9	60.4
All topics									
{Both}	51.0	59.4	52.3	49.0	56.7	60.2	55.0	57.1	**61.9**
{Male}	51.8	54.9	43.5	47.1	55.7	**56.1**	53.7	52.9	55.7
{Fem.}	50.2	**73.6**	51.7	58.5	62.6	72.8	55.1	61.9	64.9

Best results are highlighted in bold.

integrating them with other modalities does not enhance the performance.

Third, gender information clearly improves the performance. In the majority of cases, the highest accuracy achieved separately for males or females outperforms the accuracy achieved using both genders combined, despite the fact that using data from both genders yields a larger dataset of approximately double the size. Furthermore, it can be noticed that males and females have up to 81% and 64% accuracy, respectively, using linguistic features in the "Mock Crime" scenario. However, using both genders, the performance using individual and combined modalities does not exceed the baseline. This indicates that gender-based learning captures information that enhances the deception detection performance. Fourth, using "All topics" does not in general improve the results over the individual scenarios, which suggest that domain-specific information is valuable for the task of deception detection. Finally, the SVM accuracy figures are up to 66%, and showed comparable performance to decision tree, with slight improvement, in 7 out of 12 cases for "All Modalities". Interestingly, for all three individual topics and for "All topics" combined, SVM achieves better prediction rates for females as well.

Human Performance

To put our task in context, we evaluate the human ability to detect deception when the gender of the potential deceiver is known, and partial data streams are used as source of information. Since our goal is to assess the layperson ability to identify deceit, the annotation was conducted by nonexperts in deception, who judge the veracity of the information presented in different format. Specifically, we evaluate four modalities: the transcript (Text); the audio track of the video (Audio); the video with muted audio (Silent video); and audio and video played simultaneously (Full video).

Table 3. Human accuracy versus automatic deception detection (Sys) accuracy on the deception dataset over four modalities.

	Male		Female		Both	
	Human	Sys	Human	Sys	Human	Sys
Text	65.3	54.9	67.3	73.6	66.3	59.4
Audio	60.6	NA	63.3	NA	62	NA
Silent Video	50	47.1	52	58.5	51	49
Full Video	63.3	56.1	59.3	72.8	61.3	60.2

From the large dataset, we select a random gender-balanced sample of 60 subjects, consisting of 5 recordings per subject and comprising a total of 300 recordings. The recordings were divided in four sets, and each set was labeled by one human annotator. Annotations were conducted using a web-based annotation interface to label the presented modality as either "Deception" or "Truth" according to the annotator's perception of truthfulness or falsehood. During the annotation, the annotators knew the subject's gender beforehand, and the modalities were shown in the following order: first either Text or Silent video (order chosen randomly), then Audio, followed by Full video.

Table 3 shows the accuracy of the human annotators, alongside the developed system for the four modalities by gender. Results in this table show interesting trends. First, both humans and automatic classifiers improve their accuracy in deception detection with the availability of more modalities, except for text only (which presumes accurate transcriptions). Second, knowing that the potential deceiver is a female helps both the humans and the system to detect deception more accurately, thus confirming our hypothesis that deception is more easily identifiable among females. Overall, the multimodal setting benefits from gender information thus suggesting that gender-specific models can lead to the development of improved deception detection systems.

CONCLUSION

In this paper, we introduced a novel gender-balanced deception dataset and conducted an analysis using linguistic, thermal, and visual features to develop an automated, contact-free deception detection approach. We analyzed the patterns associated with the two genders as they responded truthfully/deceptively and investigated the performance of the system when it was fed with multimodal features.

Overall, our experimental results suggested that deception is easier to detect among females than males. Additionally, we showed that different, and in some cases even opposed thermal, linguistic, and visual patterns were used by males and females as they acted deceptively, such as the variation in the temperatures in certain scenarios, the length of the responses, the word classes used, and the specific gestures such as gaze, head, lips and eyebrows movements.

We also showed that multimodal deception detection models that learn from each gender separately can outperform non-gender specific models, despite having less training data available. This suggests that current approaches for deception detection can benefit from personalizing their systems toward specific demographic dimensions, instead of relying on general trends in human behavior. Finally, we compared our system accuracy against the human ability of identifying deception when the gender of the potential deceiver was known, and observed comparable performance between our system and humans when more modalities and gender information became available.

REFERENCES

1. M. Ott, Y. Choi, C. Cardie, and J. Hancock, "Finding deceptive opinion spam by any stretch of the imagination," in "*Proc. 49th Annu. Meet. Assoc. Comput. Linguistics: Hum. Lang. Technol.*, 2011, pp. 309–319.

2. H. Liu and R. Mihalcea, “Of men, women, and computers: Data-driven gender modeling for improved user interfaces,” in *Proc. Int. Conf. Weblogs Soc. Media*, 2007.
3. V. Pérez-Rosas and R. Mihalcea, “Experiments in open domain deception detection,” in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2015, pp. 1120–1125. Available at: https://aclweb.org/anthology/D/D15/D15-1133.
4. P. Tilley, J. F. George, and K. Marett, “Gender differences in deception and its detection under varying electronic media conditions,” in *Proc. 47th Hawaii Int. Conf. Syst. Sci.*, 2005, vol. 1, pp. 1–9, Available at: http://doi.ieeecomputersociety.org/10.1109/HICSS.2005.284
5. M. Abouelenien, V. Pérez-Rosas, R. Mihalcea, and M. Burzo, “Detecting deceptive behavior via integration of discriminative features from multiple modalities,” *IEEE Trans. Inf. Forensics Secur.*, vol. 12, no. 5, pp. 1042–1055, May 2017, doi: 10.1109/TIFS.2016.2639344.
6. M. Abouelenien, V. Pérez-Rosas, B. Zhao, R. Mihalcea, and M. Burzo, “Gender-based multimodal deception detection,” in *Proc. Symp. Appl. Comput.*, 2017, pp. 137–144, doi: 10.1145/ 3019612.3019644.
7. S. Feng, R. Banerjee, and Y. Choi, “Syntactic stylometry for deception detection,” in *Proc. 50th Annu. Meet. Assoc. Comput. Linguistics: Short Papers Volume 2*, 2012, pp. 171–175, Available at: http://dl.acm.org/citation.cfm?id=2390665
8. M. G. Frank and P. Ekman, “Automatic recognition of facial actions in spontaneous expressions,” *J. Personality Soc. Psychol.*, vol. 86, no. 3, pp. 486–495, Mar. 2004.
9. D. Cohen, G. Beattie, and H. Shovelton, “Nonverbal indicators of deception: How iconic gestures reveal thoughts that cannot be suppressed,” *Semiotica*, vol. 2010, no. 182, pp. 133–174, 2010.
10. M. Abouelenien, V. Pérez-Rosas, R. Mihalcea, and M. Burzo, “Deception detection using a multimodal approach,” in *Proc. 16th Int. Conf. Multimodal Interact.*, 2014, pp. 58–65.
11. J. Conrads, B. Irlenbusch, R. M. Rilke, and G. Walkowitz, “Lying and team incentives,” *J. Econ. Psychol.*, vol. 34, pp. 1–7, 2013, Available at: http://www.sciencedirect.com/science/article/pii/S0167487012001365, doi: 10.1016/j.joep.2012.10.011.
12. R. E. Guadagno, B. M. Okdie, and S. A. Kruse, “Dating deception: Gender, online dating, and exaggerated self-presentation,” *Comput. Hum. Behav.*, vol. 28, no. 2, pp. 642–647, 2012, Available at: http://www.sciencedirect.com/science/article/pii/S0747563211002548, doi: 10.1016/j.chb.2011.11.010.
13. M. Yancheva and F. Rudzicz, “Automatic detection of deception in child-produced speech using syntactic complexity features,” in *Proc. 51st Annu. Meet. Assoc. Comput. Linguistics (Volume 1: Long Papers)*, 2013, pp. 944–953, Available at: http://www.aclweb.org/anthology/P13- 1093
14. R. Mihalcea and S. Pulman, “Linguistic ethnography: Identifying dominant word classes in text,” *Computational Linguistics and Intelligent Text Processing*. New York, NY, USA: Springer, 2009, pp. 594–602.
15. X. Lu, “Automatic analysis of syntactic complexity in second language writing,” *Int. J. Corpus Linguistics*, vol. 15, no. 4, pp. 474–496, 2010.
16. J. Allwood, L. Cerrato, K. Jokinen, C. Navarretta, and P. Paggio, “The MUMIN coding scheme for the annotation of feedback, turn management and sequencing phenomena,” *Lang. Resour. Eval.*, vol. 41, no. 3–4, pp. 273–287, 2007, Available at: http://dx.doi.org/10.1007/s10579-007-9061-5, doi: 10.1007/s10579-007-9061-5.
17. T. Qin, J. Burgoon, J. Nunamaker, and J. F., “An exploratory study on promising cues in deception detection and application of decision tree,” in *Proc. 37th Annu. Hawaii Int. Conf. Syst. Sci.*, 2004, pp. 23–32.
18. T. Qin, J. K. Burgoon, J. P. Blair, and J. F. Nunamaker, “Modality effects in deception detection and applications in automatic deception-detection,” in *Proc. 38th Hawaii Int. Conf. Syst. Sci.*, 2005, pp. 23b–23b.

Mohamed Abouelenien is an Assistant Professor with the Department of Computer and Information Science, the University of Michigan-Dearborn, Dearborn, MI, USA. He was a Postdoctoral Research Fellow with Electrical Engineering and Computer Science Department, the University of Michigan, Ann Arbor from 2014–2017. His areas of interests include multimodal deception detection, multimodal sensing of thermal discomfort and drivers' alertness levels, emotion and stress analysis, machine learning, image processing, face and action recognition, and natural language processing. He received the Ph.D. degree in computer science and engineering from the University of North Texas, Denton, TX, USA, in 2013. Contact him at zmohamed@umich.edu.

Mihai Burzo is an Assistant Professor of mechanical engineering with the University of Michigan-Flint, Flint, MI, USA. Prior to joining University of Michigan, Ann Arbor, MI, USA, in 2013, he was an Assistant

Professor at University of North Texas, Denton, TX, USA. His research interests include heat transfer in microelectronics and nanostructures, thermal properties of thin films of new and existing materials, multimodal sensing of human behavior, computational modeling of forced and natural heat convection. He was the recipient of several awards, including the 2006 Harvey Rosten Award For Excellence for "outstanding work in the field of thermal analysis of electronic equipment." Contact him at mburzo@umich.edu.

Verónica Pérez-Rosas is an Assistant Research Scientist with the Computer Science and Engineering Department, the University of Michigan. Her research interests include machine learning, natural language processing, computational linguistics, affect recognition, and multimodal analysis of human behavior. Her research focuses on developing computational methods to analyze, recognize, and predict human affective responses during social interactions. She received the Ph.D. degree in computer science and engineering from the University of North Texas, Denton, TX, USA, in 2014. Contact her at vrncapr@umich.edu.

Rada Mihalcea is a Professor with the Computer Science and Engineering Department, the University of Michigan. Her research interests include computational linguistics, multimodal behavior analysis, and computational social sciences. She received the Ph.D. degree in computer science and engineering from Southern Methodist University, Dallas, TX, USA, in 2001, and the Ph.D. degree in linguistics from Oxford University, Oxford, U.K., in 2010. She was the recipient of a National Science Foundation CAREER award (2008) and a Presidential Early Career Award for Scientists and Engineers (2009). In 2013, she was made an honorary citizen of her hometown of Cluj-Napoca, Romania. Contact her at mihalcea@umich.edu.

Haitian Sun is currently working toward the Master's degree at Carnegie Mellon University, Pittsburgh, PA, USA. He received the undergraduate degree from the University of Michigan, Ann Arbor, MI, USA, in 2016. Contact him at htsun@umich.edu.

Bohan Zhao is a Software Engineer with Google, Mountain View, CA, USA. He received the Master's degree from the University of Michigan, Ann Arbor, MI, USA, in 2016. Contact him at zhaoboha@umich.edu.