

# Automatic Nonverbal Behavior Indicators of Depression and PTSD: Exploring Gender Differences

Giota Stratou, Stefan Scherer, Jonathan Gratch, Louis-Philippe Morency  
University of Southern California, Institute for Creative Technologies, Los Angeles  
Email: {stratou, scherer, gratch, morency}@ict.usc.edu

**Abstract**—In this paper, we show that gender plays an important role in the automatic assessment of psychological conditions such as depression and post-traumatic stress disorder (PTSD). We identify a directly interpretable and intuitive set of predictive indicators, selected from three general categories of nonverbal behaviors: affect, expression variability and motor variability. For the analysis, we introduce a semi-structured virtual human interview dataset which includes 53 video recorded interactions. Our experiments on automatic classification of psychological conditions show that a gender-dependent approach significantly improves the performance over a gender agnostic one.

## I. INTRODUCTION

Recent advances in the field of automatic facial feature tracking [1], [2] are revolutionizing our ability to analyze and understand nonverbal behavior, and spawning a host of novel applications. One promising use of this technology is the automatic analysis of nonverbal behaviors associated with mental illness. Extensive research in the behavioral sciences has demonstrated a link between specific psychological disorders, for example depression, and patterns of nonverbal behavior [3], [4]. Recognized these nonverbal indicators, however, often relies on the expert judgments of trained clinicians and are often not easily quantifiable [3]. Automatic detection of such indicators could assist a clinician by supporting his/her observations and by providing a more systematic measurement and quantification of nonverbal patterns both within and across clinical sessions. Additionally, fully-automated techniques might serve a pre-screening instrument for patients, complementing the self-reported questionnaires currently used for this purpose.

Many challenges confront the development of robust indicators of psychological illness. There has been some promising work to overcome those [5], [6], but there are still some limitations to address. First, there has been little work on the automatic computational analysis side that sheds light in the gender specific behaviors in illness. Most of the researchers take a gender-independent approach. There are a few exceptions [7], but even in those cases individual indicators have not been studied separately for the two genders. Second, existing indicators are often derived from extreme exemplars of the condition (e.g., severe depression) and may not generalize to more common forms of the illness. Finally, most research on automatic detection of distress focuses on depression and anxiety leaving the condition of post-traumatic stress disorder (PTSD) less covered. PTSD can cause significant impairment in social and occupational functioning [8] and it is common for war veterans but also appears in general population as well.

In this paper, we show that gender plays an important role in the automatic analysis of psychological conditions. We

introduce a semi-structured interview dataset which contains 53 dyadic interactions with participants from general population. We identify a directly interpretable and intuitive set of predictive indicators, selected from three general categories of nonverbal behaviors: affect, expression variability and motor variability. We show that a gender-dependent approach improves the results of classification for distress assessment and provides meaningful insight on gender differences for depression and PTSD.

The following section describes related work. In Section III we introduce the *Virtual Human Distress Assessment Interview Corpus* (VH DAIC) dataset. In Section IV we explain our automatic techniques for behavior extraction. We proceed with gender specific analysis of automatic indicators in Section V. In Section VI we present the classification experiments for the two distress conditions, compare a gender agnostic to a gender dependent approach and discuss the results in Section VII. Finally, Section VIII presents conclusions and future work.

## II. RELATED WORK

There has been extensive study in the field of psychology about depression characteristics. Ellgring mentions that a dysphoric state, latency in response, motor retardation (or lack of motor), lack of emotional variability (or lack of facial expressions) and hostility/aggressive behavior are central to depression [3]. Similar findings are reported by others. In particular, Troisi *et al.* [9] explored gender differences in clinical interviews with depressed patients and reported that both male and female depressed patients showed global restriction of nonverbal expressiveness, with hostility being the only behavioral category on which they scored higher than non-depressed volunteers. The authors found differences in nonverbal behavior of males and females reporting that depressed women showed more socially interactive behaviors than depressed men and that their modality of interacting included higher levels of both nonverbal hostility and submissive/affiliative behaviors. Recent work has also been focusing on particular indicators like Reed *et al.* that explore smile under positive stimuli for depressed population [10].

On the side of automatic assessment of depression there has been promising effort by Cohn *et al.* [5] achieving 79% accuracy using facial actions measured by active appearance modeling (AAM) in a population of clinically depressed patients undergoing treatment. McIntyre *et al.* presented an approach for measuring facial activity as a measure of depression by grouping face areas [6], but do not report results. One other team has taken a gender dependent approach to the automatic detection of depression: Maddage *et al.* [7] who classified depression in adolescents using Gabor wavelet

features and compared gender independent modeling approach to a gender based one, finding the latter to improve accuracy by 6%. However, their model used only adolescents, with limited population (8 participants from a clinical setup) and they do not report any analysis on the behavioral indicators of depression for the two genders. Also, previous research has shown that bodily dynamics and specifically head motions, are correlated with affective states when studied in complex learning scenarios [11], but so far most work on automatic indicators of disorder focuses on facial expressions while head motions have not been examined in that context.

In contrast with depression, PTSD has not been examined as extensively. On the clinical side, work on PTSD reports that anger/aggression are often observed in interactions of traumatized patients as well as less genuine joy [12]. PTSD and depression often co-occur (in what is known as co-morbidity) and some researchers suggest they are best viewed as reflecting a more general underlying condition known as generalized distress (e.g. see [13]). In the current article we treat PTSD and depression as distinct constructs (though we revisit this issue in the discussion section).

One of the main novelties of this paper is that we study the conditions in a general population, which is different than the other studies that use clinical cases. Also, we identify that conditions have gender-dependent affects on indicators. We extract such indicators automatically and we show that a gender-dependent approach improves performance on classification. As additional benefits of this work, we are looking at the aspect of head motion, that has not been covered in that context on the automatic side, and our work also includes analysis for PTSD that has been understudied.

### III. VIRTUAL HUMAN DISTRESS ASSESSMENT INTERVIEW CORPUS

In this section, we introduce the *Virtual Human Distress Assessment Interview Corpus* (VH DAIC) dataset, which is a general population distress assessment dataset that follows similar protocol as the Distress Assessment Interview Corpus (DAIC), described in [14]. The focus of the dataset is distress assessment of participants and it includes recordings of dyadic interactions and information about the participants' condition based on a series of pre-study questionnaires. In this dataset the participants interact with a virtual human in a Wizard-of-Oz paradigm.<sup>1</sup>

#### A. Configuration

In total the dataset includes 53 participants from general population, who were recruited using Craigslist and met some basic requirements (age, language, adequate eyesight). The participant pool covers different age, gender groups and racial backgrounds. Specifically, the participant pool breaks down to 32 men and 21 women of average age 41.2 years (std=11.6).

The interaction lasted on average about 10 minutes and it was of a question-based nature. It started with the virtual human introducing the purpose and the mode of the interaction and then asking a series of questions. During this time the participant was given time to talk in response to those questions and the virtual human was displaying listening behavior. The questions asked were mostly of general content like "what did you study at school?" and "do you have trouble sleeping?". The

TABLE I. VH DAIC POPULATION

Gender	PTSD positive	PTSD negative	Depression positive	Depression negative	Total
men	10	22	7	25	32
women	12	9	10	11	21
Total	22	31	17	36	53

virtual human's question choices, follow-ups and nonverbal behavior were controlled from a panel by two human 'wizards' situated in another room.

All participants were recorded in the same configuration, seated in front of a large screen where the virtual human was displayed. The recording devices include a web-camera (Logitech 920 720p) aiming at the participant face, a Microsoft Kinect device for Windows recording upper body video and depth data and a head-mounted microphone (Sennheiser HSP 4-EW-3) for the audio.

#### B. Psychological condition assessment

For the condition assessment the participants were asked to fill in a series of questionnaires including among others the PTSD Checklist-Civilian version (PCL-C) [15] and the Patient Health Questionnaire, depression module (PHQ-9) [16]. PHQ-9 is one of the most widely used screening instruments for depression. Although such self-administered questionnaires should not be seen as a substitute for a diagnosis by a trained clinician for decisions regarding treatment, for the present purpose (i.e., identifying individuals likely to be suffering from depression) it has been shown to have high sensitivity and specificity when compared with clinical diagnoses [17]. PCL-C is a widely used screening instrument for PTSD [18] and also shows high sensitivity and specificity for this clinical condition [19]. The dataset provides extracted scores for PTSD and depression severity, respectively, as well as a binary decision on the condition (positive or negative) based on the PCL-C and PHQ-9 standards. The database population statistics are shown in TABLE I. Comparing the scores of PTSD and depression, we observed a correlation of 0.863, so the two conditions often comorbid.

### IV. AUTOMATIC BEHAVIOR EXTRACTION

In the following subsections we will first motivate our choices of nonverbal behavior to examine and we describe our approach to extract them automatically.

#### A. Motivation

Based on a collection of various clinical observations [3], we identify three main categories of nonverbal behaviors in interactions that are indicative of distress:

**Affect.** Previous work suggests that displays of aggression and hostility are tied to both depression and PTSD [3], [12]. Displays of grief have also been traditionally linked to depression [3], [20]. There are also numerous observations that displays of joy [10] are diminished in clinically depressed population. Joy has been linked with felt happiness [4] and correlated negatively with felt grief whereas displays of anger and contempt have been found to have a positive correlation with felt grief [21].

⇒ This is a good motivation to look at the intensity of expressions of *Anger*, *Disgust*, *Contempt*, *Joy* as measures of affect, as well as a few related facial Action Units (AU).

**Emotional Variability:** The homogeneity of an affective level and the total facial activity are considered good indicators

<sup>1</sup>Sample interaction at: <http://www.youtube.com/watch?v=ejczMs6b1Q4>

TABLE II. EXAMPLE OF BEHAVIOR INDICATORS SHOWING GENDER DIFFERENCES IN TREND

	Feature	Gender	Hedge's g	p-value	trend
Depression	AU4	Men	0.51	0.237	↑
		Women	-0.92	<b>0.042</b>	↓
	Disgust	Men	-0.11	0.791	~
		Women	-0.90	<b>0.046</b>	↓
	Contempt	Men	0.04	0.930	~
		Women	0.86	0.054	↑
PTSD	AU4	Men	0.76	<b>0.050</b>	↑
		Women	-1.41	<b>0.003</b>	↓
	Disgust	Men	0.84	<b>0.031</b>	↑
		Women	-1.22	<b>0.009</b>	↓
	Contempt	Men	0.10	0.797	~
		Women	1.05	<b>0.020</b>	↑

of distress. Reduced facial behavior, also mentioned as lack of emotional variability, is considered a valid indicator for depression; and in clinical studies a 'flat, mask like face' has also been reported as indicator of depression [3].

⇒ This serves as good motivation to examine *emotional variability* as a feature, and also the intensity of a *neutral* face that can be another measure of 'emotional flatness'.

**Motor Variability** or motor retardation has also been observed in depressed population [3] including reduced hand gesturing and/or head movements. Reduced eyebrow movements is a special case of this, covered separately in emotional variability. This is a very interesting aspect of nonverbal behaviors which is usually neglected in automated analysis for distress indicators.

⇒ As a measure of motor variability we will look at the *head movement variance*.

### B. Selected Feature Extraction

Based on our observations we focus on elements of affect, emotional variability and motion variability that can be extracted automatically. More specifically, we extract the following signals:

**Basic expressions of emotion:** this group includes {*Anger, Disgust, Contempt, Fear, Joy, Surprise, Sadness, Neutral*} which are the 7 basic expressions of emotion, plus 'Neutral' face which measures lack of emotions. Most of the 7 basic expressions are individually tied to indicators in the affect category, like Joy or Anger, so measuring their intensity is valuable. Also, looking at the variance of these expressions all together, is a good measure of emotion variability as discussed above. In the same category, the intensity of the 'Neutral' expression is a good measure of emotional flatness, or lack of emotion.

**Action Units:** in the analysis we also examine a few related AU's in the general eye area: {*AU4* (brow lowerer), *AU7* (lid tightener), *AU9* (nose wrinkler)} and mouth area: {*AU12* (lip corner puller)}. *AU4* intensity is a measure of frown and it appears predominantly in the expressions of anger and fear. *AU7* intensity is a measure of eyelid tightening and can appear sometimes in anger and joy. *AU9* intensity is a measure of nose wrinkling and it appears mostly in the emotion of disgust or contempt. Finally *AU12* intensity is a measure of smiling and it appears in joy [4]. We selected these action units to support the expressions of anger, disgust, contempt and joy that we examine as indicators.

**Head Gesturing:** in this category we extract signals of head rotation in all three directions {*HeadRX* (Head rotation-Up/Down), *HeadRY* (Head rotation-side), *HeadRZ* (Head

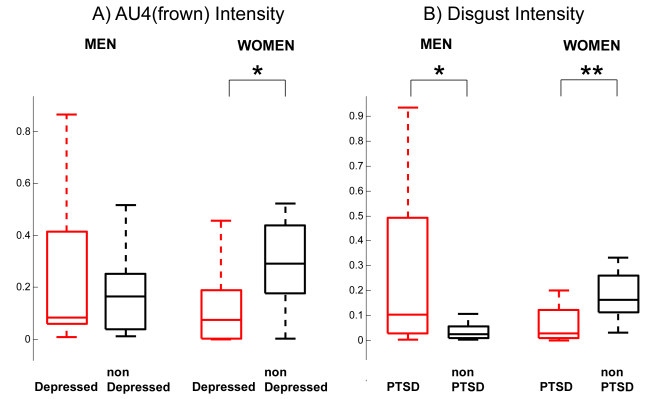


Fig. 1. Example of behavior indicators showing gender differences in trend. In both the cases of A)AU4 in depression and B)Disgust in PTSD, the conditions have opposite trends among genders. Statistically significant differences( $p \leq 0.5$ ) are shown with a star(\*)

tilt)}. From these signals we can extract information about the head gaze and the head rotation variability.

At this point we would like to mention that the list of extracted features is not exhaustive, and especially in the AU group where one can find additional wealth of information about expressivity and affect. We extracted this specific pool of features to showcase particular examples of indicators based on our previous observations. Our exploration included additional AUs in the mouth area, some of them linked to depression by previous literature [5], however concerns of noise by mouth movement due to speech, led us to explore further and report in future work.

### C. System For Automatic Sensing

In this paper we investigated nonverbal indicators of depression and PTSD using visual cues extracted automatically from the web-camera video aimed at the participant face. For the analysis of the participant videos we apply our multimodal sensing framework, called MultiSense, that has integrated several tracking technologies. The benefit of such a system is that the multiple technologies can run in parallel in a synchronized manner allowing for inter-module cooperation for performance improvement and information fusion. Our sensing system provides 3D head position-orientation, facial tracking based on GAVAM HeadTracker [22] and CLM-Z FaceTracker [2] and basic emotion analysis based on SHORE Face Detector [23]. In this analysis we also added results from the Computer Expression Recognition Toolbox (CERT) [1] for expression recognition and facial Action Unit (AU) scores. When available, we used our system's confidence report on the output to automatically screen out bad frames when analyzing the signals. In the next section we explore how discriminative these indicators are for the conditions of depression and PTSD.

## V. ANALYSIS OF INDICATORS AND GENDER DIFFERENCES

In this section we analyze the automatically extracted behavior indicators with the following goals: i) to identify indicators correlated with depression and PTSD, and ii) to study the effect of gender on these indicators. This study will inform our next set of experiments which focuses on depression and PTSD classification. In the following subsections, we first explain our statistical analysis and then showcase the

TABLE III. EXAMPLE OF BEHAVIOR INDICATORS SHOWING GENDER SIMILARITIES IN TREND

	Feature	Gender	Hedge's g	p-value	trend
Depression	HeadRot Variation	Men	-0.57	0.185	↓
		Women	-0.89	<b>0.047</b>	↓
	Emotion Variation	Men	-0.59	0.166	↓
		Women	-0.67	0.128	↓
PTSD	HeadRot Variation	Men	-0.74	0.054	↓
		Women	-0.59	0.175	↓
	Emotion Variation	Men	-0.39	0.299	↓
		Women	-0.89	<b>0.045</b>	↓

differences and similarities of our indicators when used to describe depression and PTSD.

### A. Statistical Analysis

Our goal is to examine the effect of the psychological conditions on our behavioral indicators, to study the differences and similarities between genders. As a measure of effect size we use 'Hedge's g' [24], a descriptive statistic that conveys the estimated strength of an effect by estimating how many standard deviations separate the two distribution means. For the purposes of this analysis we call the direction of that effect a *trend*. We consider a Hedge's  $g \leq -0.4$  to show existence of at least moderate effect with negative trend (↓). A psychological condition showing effect with negative trend means that the depressed (or PTSD-afflicted) population showed lower intensity on that indicator. Symmetrically, an indicator with Hedge's  $g \geq 0.4$  means that the psychological condition has an effect on the indicator with positive trend (↑). Effect sizes of smaller absolute value than 0.2 are considered to show negligible effect (~). We also report the t-test statistical significance 'p' of the difference of the distributions between distressed and non-distressed participants, to complement the Hedge's g effect size.

### B. Differences In Gender Trends

We start our analysis by focusing on trend differences between genders. Specifically, we identify two types of such indicators: 1) the first type describes indicators where the psychological condition has an opposite trends for the two genders (i.e. there is a gender-dependent crossover interaction). For example, the condition having a negative effect for men and positive for women (↓,↑) will be categorized as first type, and 2) the second type describes indicators where the condition has effect only on one gender and negligible effect (~) on the other gender. This category could include an indicator where the condition shows a positive trend for men, but no trend (no effect) for women (↑,~).

TABLE II shows indicators for both the conditions of depression and PTSD, with gender differences in trends and the effect sizes of those trends. We see that for frowning (AU4) both psychological conditions have a statistically significant effect on the frowning intensity, for both genders. More interestingly, the trends for men and women are going in opposite directions (first type we described). Specifically, as seen in Figure 1A, depressed men tend to display more frowning than the non-depressed men, whereas women display more frowning when they are non-depressed. Another interesting indicator is Disgust for PTSD, also shown in Figure 1B. It shows that PTSD-afflicted men tend to display more disgust than non-afflicted men while women display more when they are non-afflicted than the PTSD-afflicted ones. TABLE II, also

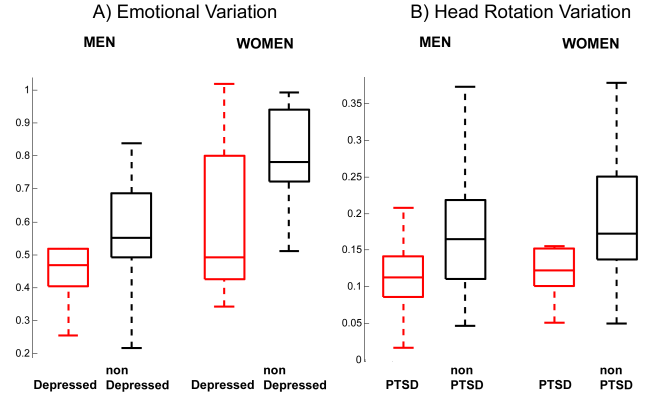


Fig. 2. Example of behavior indicators showing gender similarities in trend. In both the cases of A) Emotional Variation in depression and B) Head Rotation Variation in PTSD, the conditions show same trends among genders.

shows two cases where the condition has an effect only for one gender: Contempt for PTSD and depression and Disgust for depression, Contempt in particular seems to be significantly discriminative for women, with a positive trend, but not at all informative for men. It is an interesting indicator because it is the only 'negative' expression from our set of behavior indicators that distressed women seem to express more than non-distressed ones.

### C. Similarities in gender trends

We also identified indicators where the psychological condition has an effect with similar trends for both genders. These cases show negative trend (↓,↓) or positive trend (↑,↑) for both genders. TABLE III summarizes the effect size of such indicators. It is interesting to observe that for the indicator of Head Rotation Variance both the conditions show a negative trend for both men and women and for both the psychological conditions. The case of PTSD can be seen in Figure 2B. Similarly, the Emotional Variance is discriminative with negative trend for both genders and for both the psychological conditions. The distributions for depression are shown in Figure 2A).

We would like to point out that even though the same trend is observed for both genders, these indicators can still show gender-dependent differences. A good example is depicted in Figure 2A where the gender has an effect on the Emotional Variance indicator. Women over all, in both distressed and non-distressed conditions seem to showcase more emotional variability than men. All these observations serve as a good indication that a gender-dependent approach will benefit the assessment of depression and PTSD.

## VI. CLASSIFICATION EXPERIMENTS FOR DEPRESSION AND PTSD

In this section we test the discriminative power of our behavior indicators for the conditions of depression and PTSD by using them as features in a classification experiment. Our experimental hypothesis is that separating the two genders in a gender-dependent manner improves performance. We base this hypothesis on the observed trends (sometimes in the opposite direction) from the statistical analysis described in the previous section. As a result we are expecting that the discriminative power of these indicators may increase when separating the two genders. In the following sub-sections we

TABLE IV. CLASSIFICATION RESULTS FOR DEPRESSION AND PTSD. WE SHOW THAT A GENDER-DEPENDENT MODEL PERFORMS BETTER THAN A GENDER-INDEPENDENT ONE.

	Population	Baseline	Gender-Independent	Gender-Dependent
		F1	F1	F1
<b>Depression</b>	All	0.576	0.722	-
	Men	0.610	0.756	<b>0.808</b>
	Women	0.512	0.682	<b>0.858</b>
<b>PTSD</b>	All	0.540	0.785	-
	Men	0.579	0.739	<b>0.811</b>
	Women	0.533	0.840	<b>0.908</b>

describe the compared models, the methodology we follow for the classification experiment and present our results.

#### A. Models

In the experiments we evaluate the performance of 3 models: **Baseline** which uses the majority vote where all observations are given the same predicted label, **Gender-Independent** which is one trained model on the whole population (both genders), and **Gender-Dependent** which separates two separate models, trained on separate genders.

In order to be able to compare performance by gender, we tested separately both approaches on the two groups of 'Men' and 'Women'. The Gender-Dependent models are tested on their respective genders. Our goal is to identify differences in performance that arise from the separation of the two genders. For the Baseline and the Gender-Independent model, we also test on the whole population.

#### B. Feature Representation

Using the automatic sensing framework described in Section IV.C we extracted the behavioral signals and computed basic summary statistics for each interaction in our dataset (Section III). We use the average and the standard deviation of a signal over the whole interaction as measures of variation of the behavioral signal over the whole interaction.

In the case of the Action Units we also introduced a positive thresholded signal in order to take into account only the frames where the AU was found active. The Emotional Variation was computed by aggregating the variances of the 7 emotions mentioned in Section IV.B. For the Head Rotation Variance we added up the variance of the head rotation in all 3 axes. We also introduced a feature that combines the effect of the three 'eye-narrowing' related action units (AU4-AU7-AU9). Our final feature pool contained 20 features.

#### C. Classification

As a simple approach, we chose a Naive Bayes classifier [25] which has the advantage of having a limited number of hyperparameters. For our experiments we performed a Leave-One-Participant-Out testing and greedy forward feature selection. This experimental methodology was designed to show user-independent results. Each classifier contained two classes: PTSD vs. non-PTSD or depressed vs. non-depressed. As a measure of performance we are using F1 score which is the harmonic average of precision and recall (averaged for both labels).

#### D. Results

In TABLE IV we show our classification results. The table compares the results of the gender-independent approach

(Gender-Independent) with our gender-dependent approach (Gender-Dependent) where we train separate models for men and women. Results show that the gender-dependent approach performs better for both test groups of 'Men' and 'Women'. Also, the gender-independent approach performs better than the baseline for all test groups.

## VII. DISCUSSION

Our classification results confirm the trends shown in our statistical analysis. Specifically, we showed that separating men and women when assessing their nonverbal behaviors improves the performance of classification. Our gender-dependent classification can take full advantage of behavior indicators, such as disgust in PTSD and frowning (AU4) in PTSD and in depression. These indicators showed opposite trends for men and women. Moreover, the indicators that show trend for only one gender and don't affect the other, may lose their discriminative power in a gender-independent classification, or wrongfully transfer their discriminative effect into the other gender.

Our results reflect findings in clinical and social studies that support the claim that men and women demonstrate different nonverbal behaviors when depressed [9]. There are intrinsic differences in nonverbal behaviors among genders [26], sometimes amplified or attenuated by social norms and gender-related expectations [27]. However, one should be cautious about the interpretation of such phenomena. For one, elicited behaviors are often influenced by the interaction style [28] and the lack of or plethora of stimuli. Secondly, on the automatic part of the feature extraction, one should take into consideration the possibility of tracker gender bias when designing the indicators.

The interaction style becomes a very important factor to control, since parameters like the gender of the interviewer or, in our case, the interviewer being a virtual agent, can affect the genders' perception [29]. In addition to the above, some psychological conditions like depression and PTSD have different base rates among the two genders [30], thus making it difficult to produce balanced populations for studies, and this could be seen as an additional motivation why gender-dependent analysis might be beneficial.

At this point, we would like to mention that the introduced gender-dependent approach does not hinder nor discourage a fully automatic approach for producing indicators for depression and PTSD. Gender recognition can be performed automatically. As a proof of concept, we evaluated the performance of our system's real-time gender detection (based on SHORE Face Detector [23]). By taking the first 3sec of the interactions our system correctly classified 84% of the participant genders. This number could be improved if we add audio information.

The analysis in this article treated PTSD and depression as distinct clinical conditions, though it should be noted that both conditions frequently co-occur. Indeed, our sample showed similar rates of comorbidity to what has been reported in other studies. Some researchers have gone as far as to argue that PTSD and depression are simply manifestations of the same underlying disorder and that it is not meaningful to distinguish the two conditions [13]. Others argue that it is highly meaningful to differentially distinguish between these conditions and recommend distinct treatments depending on whether one or both conditions are present [31]. It would be straightforward to extend our methods to distinguishing

between different conditions and it might be possible to find nonverbal behaviors that help differentiate "pure" vs. comorbid participants (e.g., participants suffering only depression vs. those comorbid for both depression and PTSD). Our current sample size precluded such an analysis but, given sufficient data, this would be useful direction to explore.

## VIII. CONCLUSION

We identified a directly interpretable and intuitive set of automatically extracted indicators for depression and PTSD. This set includes the quantitative analysis of three general categories, namely affect, expression variability, and motor variability, and ties to the predominantly manually assessed observations within the field of clinical psychology. Moreover, we show that a gender-dependent analysis of nonverbal indicators allows for deeper insights into typical behaviors, which would otherwise be obscured within a gender-independent analysis by interactional effects between the psychological condition and gender. Our experiments revealed that gender-dependent models outperform gender agnostic approaches and improve results for both investigated psychological conditions. In the future, we plan to explore indicators based on dynamic and multimodal observations by incorporating additional modalities, such as audio, body gestures/posture as well as context/lexical patterns.

## ACKNOWLEDGMENT

This work is supported by DARPA under contract (W911NF-04-D-0005) and U.S. Army Research, Development, and Engineering Command. The content does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

## REFERENCES

- [1] G. Littlewort, J. Whitehill, T. Wu, I. Fasel, M. Frank, J. Movellan, and M. Bartlett, "The computer expression recognition toolbox (cert)," in *FG 2011, 2011 IEEE International Conference on*, 2011.
- [2] T. Baltrusaitis, P. Robinson, and L. Morency, "3d constrained local model for rigid and non-rigid facial tracking," in *CVPR, 2012 IEEE Conference*, 2012.
- [3] H. Ellgring, *Nonverbal communication in depression*. Cambridge: Cambridge University Press, 1989.
- [4] P. Ekman, D. Matsumoto, and W. V. Friesen, "Facial expression in affective disorders," *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)*, vol. 2, pp. 331–342, 1997.
- [5] J. Cohn, T. Kruez, I. Matthews, Y. Yang, M. H. Nguyen, M. Padilla, F. Zhou, and F. De la Torre, "Detecting depression from facial actions and vocal prosody," in *ACII 2009. 3rd International Conference on*, 2009, pp. 1–7.
- [6] G. McIntyre, R. Gocke, M. Hyett, M. Green, and M. Breakspear, "An approach for automatically measuring facial activity in depressed subjects," in *ACII 2009. 3rd International Conference on*, 2009.
- [7] M. Maddage, R. Senaratne, L.-S. Low, M. Lech, and N. Allen, "Video-based detection of the clinical depression in adolescents," in *EMBC 2009. Annual International Conference of the IEEE*, 2009, pp. 3723–3726.
- [8] *Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR Fourth Edition (Text Revision)*, 4th ed. Amer Psychiatric Pub, Jul. 2000.
- [9] A. Troisi and A. Moles, "Gender differences in depression: an ethological study of nonverbal behavior during interviews," *Journal of Psychiatric Research*, vol. 33, no. 3, pp. 243 – 250, 1999.
- [10] L. I. Reed, M. A. Sayette, and J. F. Cohn, "Impact of Depression on Response to Comedy: A Dynamic Facial Coding Analysis," *Journal of Abnormal Psychology*, vol. 116, pp. 804–809, 2007.
- [11] S. D'Mello and A. Graesser, "Dynamics of affective states during complex learning," *Learning and Instruction*, vol. 22, no. 2, pp. 145 – 157, 2012.
- [12] A. Kirsch and S. Brunnhuber, "Facial expression and experience of emotions in psychodynamic interviews with patients with ptsd in comparison to healthy subjects," *Psychopathology*, vol. 40, no. 5, pp. 296–302, 2007.
- [13] J. A. Russell and L. F. Barrett, "Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant," *Journal of Personality and Social Psychology*, vol. 76, no. 5, pp. 805–819, May 1999.
- [14] S. Scherer, G. Stratou, M. Mahmoud, J. Boberg, J. Gratch, A. Rizzo, and L.-P. Morency, "Automatic behavior descriptors for psychological disorder analysis," in *accepted for publication at IEEE Conference on Automatic Face and Gesture Recognition*, 2013.
- [15] E. B. Blanchard, J. Jones-Alexander, T. Buckley, and C. Forneris, "Psychometric properties of the ptsd checklist (pcl)," *Behaviour Research and Therapy*, vol. 34, no. 8, pp. 669–673, 1996.
- [16] K. Kroenke and R. L. Spitzer, "The phq-9: A new depression and diagnostic severity measure," *Psychiatric Annals*, vol. 32, pp. 509–521, 2002.
- [17] K. Kroenke, R. L. Spitzer, and J. B. W. Williams, "The phq-9," *Journal of General Internal Medicine*, vol. 16, no. 9, pp. 606–613, 2001.
- [18] E. E. Bolton, M. J. Gray, and B. T. Litz, "A cross-lagged analysis of the relationship between symptoms of ptsd and retrospective reports of exposure," *Journal of Anxiety Disorders*, vol. 20, no. 7, pp. 877–895, 2006.
- [19] M. A. Andrykowski, M. J. Cordova, J. L. Studts, and T. W. Miller, "Posttraumatic stress disorder after treatment for breast cancer: Prevalence of diagnosis and use of the ptsd checklistcivilian version (pclc) as a screening instrument," *Journal of consulting and clinical psychology*, vol. 66, no. 3, pp. 586–590, 1998.
- [20] P. Ekman and E. L. Rosenberg, *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS) (Series in Affective Science)*. New York: Oxford UP, USA, 2005.
- [21] G. A. Bonanno and D. Keltner, "Facial expressions of emotion and the course of conjugal bereavement," *Journal of Abnormal Psychology*, vol. 106, pp. 126–137, 1997.
- [22] L. Morency, J. Whitehill, and J. Movellan, "Generalized adaptive view-based appearance model: Integrated framework for monocular head pose estimation," in *FG '08. 8th IEEE International Conference on*, 2008.
- [23] C. Kublbeck and A. Ernst, "Face detection and tracking in video sequences using the modifiedcensus transformation," *Image and Vision Computing*, vol. 24, no. 6, pp. 564 – 572, 2006.
- [24] L. V. Hedges, "Distribution theory for glass's estimator of effect size and related estimators," *Journal of Educational and Behavioral Statistics*, vol. 6, no. 2, pp. 107–128, 1981.
- [25] [Online]. Available: <http://www.mathworks.com/products/statistics>
- [26] F. M. Deutsch, D. LeBaron, and M. M. Fryer, "What is in a smile?" *Psychology of Women Quarterly*, vol. 11, no. 3, pp. 341–352, 1987.
- [27] J. A. R. E. K. U. Hess, R. B., "Who may frown and who should smile? dominance, affiliation, and the display of happiness and anger," 2005.
- [28] N. C. Krämer, L. Hoffmann, and S. Kopp, "Know your users! empirical results for tailoring an agent's nonverbal behavior to different user groups," in *Proceedings of IVA*, ser. IVA'10, 2010, pp. 468–474.
- [29] P. Kulms, N. Kramer, J. Gratch, and S.-H. Kang, "It's in their eyes: A study on female and male virtual humans gaze," in *IVA*, ser. Lecture Notes in Computer Science, 2011, vol. 6895, pp. 80–92.
- [30] W. Rief, A. Nanke, A. Klaiberg, and E. Braehler, "Base rates for panic and depression according to the brief patient health questionnaire: a population-based study," *Journal of Affective Disorders*, vol. 82, no. 2, pp. 271 – 276, 2004.
- [31] D. Campbell, B. Felker, C.-F. Liu, E. Yano, J. Kirchner, D. Chan, L. Rubenstein, and E. Chaney, "Prevalence of depressionptsd comorbidity: Implications for clinical practice guidelines and primary care-based interventions," *Journal of General Internal Medicine*, vol. 22, no. 6, pp. 711–718, 2007.