J Multimodal User Interfaces (2015) 9:17–29
DOI 10.1007/s12193-014-0161-4

ORIGINAL PAPER

Automatic nonverbal behavior indicators of depression and PTSD: the effect of gender

Giota Stratou · Stefan Scherer · Jonathan Gratch · Louis-Philippe Morency

Received: 23 January 2014 / Accepted: 9 May 2014 / Published online: 8 June 2014

Abstract Recently there has been arising interest in automatically recognizing nonverbal behaviors that are linked with psychological conditions. Work in this direction has shown great potential for cases such as depression and post-traumatic stress disorder (PTSD), however most of the times gender differences have not been explored. In this paper, we show that gender plays an important role in the automatic assessment of psychological conditions such as depression and 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 employ 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.

Keywords Automatic analysis · Depression · PTSD · Gender differences · Distress indicators · Virtual human interaction

G. Stratou (✉) · S. Scherer · J. Gratch · L.-P. Morency
Institute for Creative Technologies, University of Southern California, Los Angeles, California
e-mail: stratou@ict.usc.edu

S. Scherer
e-mail: scherer@ict.usc.edu

J. Gratch
e-mail: gratch@ict.usc.edu

L.-P. Morency
e-mail: morency@ict.usc.edu

1 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]. Recognizing these nonverbal indicators, however, often relies on the expert judgments of trained clinicians and are often not easily quantifiable [4]. 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 as 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 being 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]; it is common for war veterans but 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 employ 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 Sect. 3 we introduce the *Virtual Human Distress Assessment Interview Corpus* (VH DAIC) dataset. In Sect. 4 we explain our automatic techniques for behavior extraction. We proceed with gender specific analysis of automatic indicators in Sect. 5, where we study gender differences for the two conditions of depression and PTSD. In Sect. 6, we present the classification experiments for the two distress conditions, compare a gender agnostic to a gender-dependent approach and discuss the results in Sect. 7. Finally, Sect. 8 presents conclusions and future work.

2 Related work

In this section we describe the two psychological conditions that we study, namely depression and PTSD. We refer to previous work on nonverbal behaviors that have been associated with these two conditions. Furthermore, we look at reports on gender differences in nonverbal behaviors during illness and recent work on automatic assessment of depression and PTSD.

2.1 Definitions for depression and PTSD

2.1.1 Depression

It is one of the most commonly used words in psychiatry, and it is also one of the most ambiguous. As a symptom it can mean sadness, but as a diagnosis it can be applied to people who deny feeling sad. Major depressive disorder (MDD) (also known as clinical depression, major depression, unipolar depression, or unipolar disorder) represents the classic condition in the group of depressive disorders. It is a mental disorder characterized by a pervasive and persistent low mood that is accompanied by low self-esteem and by a loss of interest or pleasure in normally enjoyable activities. This cluster of symptoms was named, described and classified as one of the mood disorders in the 1980 edition of the American Psychiatric Association's diagnostic manual. It often involves clear-cut changes in affect, cognition and neurovegetative functions. A diagnosis based on a single episode is possible, although the disorder is a recurrent one in the majority of cases. Careful consideration is given to the distinction of normal sadness and grief from a major depressive episode [8,9].

2.1.2 PTSD

It is a trauma-related disorder in which exposure to a traumatic or stressful event is listed explicitly as a diagnostic criterion. In some cases, symptoms can be well understood within an anxiety- or fear-based context; it is clear, however, that many individuals who have been exposed to a traumatic or stressful event exhibit a phenotype in which, rather than anxiety- or fear-based symptoms, the most prominent clinical characteristics are anhedonic (showcasing lack of pleasure or of the capacity to experience it) and dysphoric symptoms (showcasing unhappiness or uneasiness), externalizing angry and aggressive symptoms, or dissociative symptoms [9]. The clinical presentation of PTSD varies. In some individuals, fear-based re-experiencing, emotional, and behavioral symptoms may predominate. In others, anhedonic or dysphoric mood states and negative cognitions may be most distressing. In some other individuals, arousal and reactive-externalizing symptoms are prominent, while in others, dissociative symptoms predominate. Finally, some individuals exhibit combinations of these symptom patterns [9].

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 [10]). In the current article we treat PTSD and depression as distinct constructs (though we revisit this issue in the discussion section).

2.2 Nonverbal behaviors in depression and PTSD

There has been extensive study in the field of psychology about depression characteristics. Ellgring mentions that a dysphoric state (showcasing unhappiness or uneasiness), 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 [4]. Similar findings are reported by others. Recent work has also been focusing on particular indicators like Reed et al. that explore smile under positive stimuli for depressed population [11]. These observations fit well into some of the behavioral diagnostic criteria of depression, as mentioned by the American Psychiatric Association's diagnostic manual [9] :

- Depressed mood (feels sad, hopeless, appears tearful). This can be irritable mood for children and adolescents.
- Significant dimish in interest or pleasure.

– Psychomotor agitation or retardation (feeling of restlessness or being slowed down).
– Fatigue or loss of energy.
– Diminished ability to think or concentrate, or indecisiveness.

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]. Those fit well into some of the main diagnostic criteria of PTSD, as mentioned by the American Psychiatric Association's diagnostic manual [9]:

– Persistent negative emotional state (e.g. fear, horror, anger, guilt or shame).
– Diminish in interest/participation in significant activities.
– Feelings of detachment or estrangement from others.
– Persistent inability to experience positive emotions (e.g. inability to experience happiness, satisfaction or loving feelings).
– Irritable behavior and angry outbursts towards people or objects.
– Problems with concentration.

It is very important to note, that human clinitians judge those behaviors within context and under cultural and social norms in order to make more accurate assessments. For example, experiencing sadness because of loss is a different case than being chronically depressed; different cultures showcase different baselines of "normal" behavior, and genders follow different social norms of "accepted behaviors".

2.3 Gender differences in behaviors associated with depression and PTSD

Gender differences in depression have been studied before in clinical psychology. In particular, Troisi et al. [13] 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. They 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.

Gender differences in patients with PTSD have been studied in clinical psychology [14,15] but are most likely to focus on the risk factor of the disease among genders, the index trauma as cause of the disorder and the comorbidity rates with other disorders. Specifically, it has been reported that men are more likely to have comorbid substance use disorders and women are more likely to have comorbid mood and anxiety disorders, although many disorders comorbid with PTSD are commonly seen in both men and women. There are a few articles reporting gender differences in the presentation of PTSD such as: women are more likely to have symptoms of numbing and avoidance and men are more likely to have the associated features of irritability and impulsiveness [16], but overall behavioral differences have been understudied. One of the main reasons is that data on gender differences in PTSD among psychiatric patients are scarce because most clinical studies of PTSD have focused on patients seeking treatment for a specific trauma that is predominantly gender related, such as combat or sexual trauma [14].

2.4 Automatic assessment of depression and PTSD

On the side of automatic assessment of depression there has been recent work [17,18] that focuses on the automatic diagnosis of depression from multimodal features, exploring dynamics of the face. While the results are very promising, feature representations used in most cases do not offer much intuition in the condition. On that front 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. [6] also presented an approach for measuring facial activity as a measure of depression by grouping face areas, but do not report results. Another recent article stands out that presents evidence from manual and automatic analysis of behaviors associated with increasing severity of depression and links with the Social Withdrawal hypothesis [19]. The edge that this work brings is that it links measured behaviors under depression to a theory on inherent motivation behind those expressions. Also, previous research has shown that bodily dynamics and specifically head motions, are correlated with affective states when studied in complex learning scenarios [20]. Recent work has also shown that relative body part movements is a useful feature for automatic depression classification [21]. So far head motions had not been examined in that context with the recent exception of Girard et al. [19] who report that diminished head motion (in amplitude and velocity) is observed with severe cases of depression. 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.

One of the main novelties of this paper is that we study the conditions in participants recruited by general population, which is different than the other studies that use clinical cases. Also, we identify that conditions have gender-dependent effects 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 only recently been covered in that context on the automatic side, and our work also includes analysis for PTSD that has been understudied.

3 Virtual human distress assessment interview corpus

In this section, we will describe 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 [22]. The focus of the dataset is distress assessment of participants and it includes recordings of multimodal 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.[1] VH DAIC is a multimodal dataset including audio, video and depth recordings of the interaction.

3.1 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). By experimental design, the study sample was biased towards participants who have been diagnosed at some point in their life with depression or PTSD. The participant pool covers different age, gender groups and racial backgrounds. Specifically, the participant pool breaks down to 32 males and 21 females of average age 41.2 years (std = 11.6).

All participants were recorded in the same configuration, seated in front of a large screen where the virtual human was displayed. Figure 1 displays the screenshots of the participant and Ellie (the virtual human) placed side-by-side. 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. All the streams are recorded in a synchronized manner using the same recording platform.

[1] Sample interaction between the virtual agent and a human actor can be seen here: http://www.youtube.com/watch?v=ejczMs6b1Q4.

Fig. 1 Screenshot of one of the VH DAIC participants (*left*) interacting with the virtual human (*right*).The interaction was multimodal as it included both verbal (dialog) and nonverbal interaction (e.g. head nods, smiles)

3.2 Interaction

The interaction lasted on average about 10 min 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?", "when was the last time you felt really happy?" and "do you have trouble sleeping?". The virtual human's question choices, follow-ups and nonverbal behavior were controlled from a panel by two human 'wizards' situated in another room.

The interaction was semi-structured in the sense that it always happened in a question-based manner and the wizards were mostly following a specific sequence of questions, but adjusted the follow-up questions (like "Tell me more about that" or "Why?") and feedback (laughter or expressions of empathy like "oh!") to each individual conversation based on what was being said.

The aim was for the participants to express themselves naturally and capture all the multiple modalities of their behavior so the interaction itself was multimodal. Both audio and video of the participant were streamed live in the wizard room, and the human wizards were also responding multimodally via the virtual human by controlling both the verbal and nonverbal behaviors. Specifically, the set of nonverbal behaviors that the wizards controlled included but was not limited to facial expressions (e.g. smile, frown), head gestures (e.g. nod, tilt) and body gestures (e.g. lean forward).

Naturally, the interaction with the virtual human, being a controlled one, yields stimuli with controlled variability that depend on the complexity of the controls. Although, generally this could be seen as a constraint, we would argue that in an experimental dyadic interaction setup it can also be beneficial by standardizing the set of stimuli the participant receives, something that is not easily maintainable with a human interviewer.

3.3 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) [23] and the Patient Health Questionnaire, depression module (PHQ-9) [24]. PHQ-9 is the nine item self-report scale based directly on the diagnostic criteria for major depressive disorder in the Diagnostic and Statistical Manual Fourth Edition (DSM-IV) [8]. 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 (88 %) when compared with clinical diagnoses [25]. PCL-C is a widely used screening instrument for PTSD [26,27] , based also on the DSM-IV and shows high sensitivity and specificity for this clinical condition (0.82 and 0.83, respectively for detecting DSM PTSD diagnoses) [28]. 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 1. For the notation on the table, and moving forward in this study we would like to clarify that the characterization PTSD-positive, -negative and Depression-positive, -negative are based on the questionnaire assessments as mentioned above. Comparing the scores of PTSD and depression, we observed a correlation of 0.863, so the two conditions often comorbid. This practically means that the group of participants labeled as PTSD-positive (22 people) intersects in big part with the group of participants labeled as Depression-positive (17 people).

4 Automatic behavior extraction

In the following subsections we will first motivate our choices of nonverbal behavior to examine and then we will describe our approach to extract them automatically. We would like to mention that even though in this article we focus on modalities extracted by the video stream, VH DAIC is a multimodal dataset and parallel efforts are exploring automatic indicators in other behavioral channels such as the voice [29] and verbal [30].

4.1 Motivation

Based on a collection of various clinical observations [4] and the DSM diagnistic criteria for the conditions, we identify three main categories of nonverbal behaviors in interactions that are indicative of distress:

4.1.1 Affect

Previous work suggests that displays of aggression and hostility are tied to both depression and PTSD [4,12]. Displays of grief have also been traditionally linked to depression [4,31]. There are also numerous observations that displays of joy [11] are diminished in clinically depressed population. Joy is an expression correlated positively with self-reports of felt happiness [3] and correlated negatively with felt grief [32]. One of the diagnistic criteria for PTSD is inability to experience happiness. On the other hand, displays of anger and contempt have been found to have a positive correlation with felt grief [32] and irritability and anger outbursts have been seen within PTSD symptoms.

⇒ 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 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 [4]. One could also consider emotional expressivity as a measure of interactive behavior, lack of which can be seen in depression and PTSD (diminishing interest/participation in significant activities such as social interaction).

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

Motor Variability or motor retardation has also been observed in depressed population [4] 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 analy-

Table 1 VH DAIC population

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

sis for distress indicators. It is a special category mentioned directly as one of the diagnostic criteria for depression and can be also linked to loss of energy which is another symptom.

⇒ As a measure of intra-subject motor variability we will look at the *head movement variance* during the interaction.

4.2 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 [3]. We selected these AUs 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 tilt)}. From these signals we can extract information about the head gaze and the head rotation variability of a participant during the interaction.

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.

4.3 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 a 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 [33] and CLM-Z FaceTracker [1] and basic emotion analysis based on SHORE Face Detector [34]. In this analysis we also added results from the Computer Expression Recognition Toolbox (CERT) [2] for expression recognition and facial AU scores. MultiSense is a sensing platform employed to quantify nonverbal behaviors and it has been developed independently of this scenario. 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.

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

5.1 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' [35], 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

Table 2 Example of behavior indicators showing gender differences in trend

	Feature	Gender	Hedge's g	*p* value	trend
Depression	AU4	Males	0.51	0.237	↑
		Females	−0.92	**0.042**	↓
	Disgust	Males	−0.11	0.791	∼
		Females	−0.90	**0.046**	↓
	Contempt	Males	0.04	0.930	∼
		Females	0.86	**0.054**	↑
PTSD	AU4	Males	0.76	**0.050**	↑
		Females	−1.41	**0.003**	↓
	Disgust	Males	0.84	**0.031**	↑
		Females	−1.22	**0.009**	↓
	Contempt	Males	0.10	0.797	∼
		Females	1.05	**0.020**	↑

Bold values indicate p $\leq$ 0.05

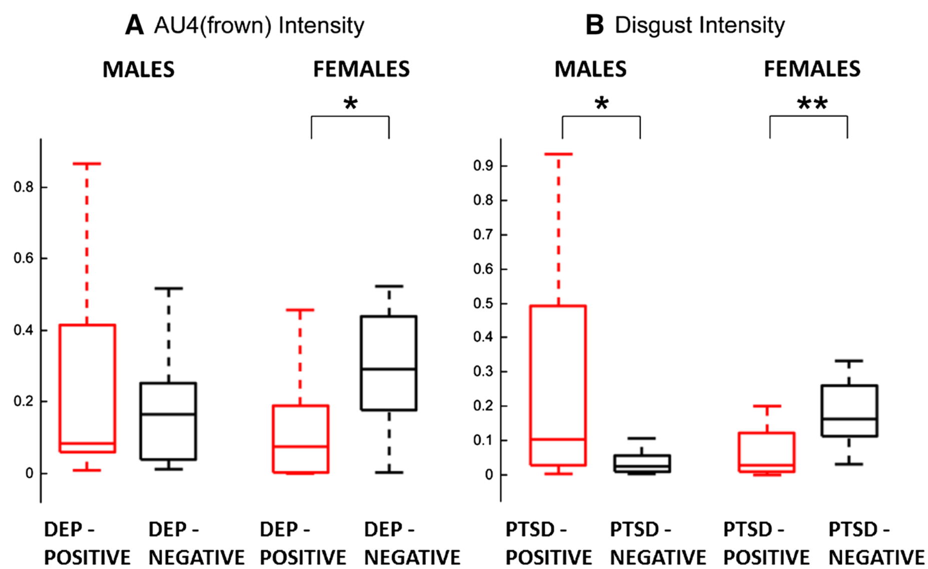

Fig. 2 Example of behavior indicators showing gender differences in trend. In both the cases of **a** AU4 in depression and **b** Disgust in PTSD, the conditions have opposite trends among genders. Statistically significant differences ($p \leq 0.5$) are shown with a *asterisk*

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.

5.2 Indicators with 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 opposite trends for the two genders (i.e. there is a gender-dependent crossover interaction). For example, the condition having a negative effect for males and positive for females (↓,↑) 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 males, but no trend(no effect) for females (↑,∼).

Table 2 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 males and females are going in opposite directions (first type we described). Specifically, as seen in Fig. 2a, depressed males tend to display more frowning than the non-depressed males, whereas females display more frowning when they are non-depressed. Another interesting indicator is Disgust for PTSD, also shown in Fig. 2b. It shows that PTSD-afflicted men tend to display more disgust than non-afflicted males while females display more when they are non-afflicted than the PTSD-afflicted ones. Table 2, also 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 females, with a positive trend, but not at all informative for males. It is an interesting indicator because it is the only 'negative' expression from our set of behavior indicators that distressed females seem to express more than non-distressed ones.

Additionally we report that anger intensity seemed to have a positive discriminative effect for males in the case of PTSD, average horizontal head direction seemed to have a positive diagnostic effect on males in depression (depressed males are significantly facing more downwards), expressions of Joy and measured AU12 seemed to be diagnostic with negative effect in the case of males in PTSD (PTSD afflicted males show less average smile intensity) but remain weak indicators for the case of females in our dataset. Finally, AU7 and AU9 show similar effect as AU4 (increasing for males in distress, decreasing for females) and that effect is strong in PTSD.

5.3 Indicators with 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 3 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 males and females and for both the psychological conditions. The case of PTSD can be seen in Fig. 3b. 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 Fig. 3a. Additionally, we report that a few other head gesture related indicators showed common trend between the two genders, such as lateral head rotation amplitude and standard deviation, horizontal and tilt head rotation standard deviation. In all of these cases participants with depression and PTSD showcase less variability in the head rotation in all directions, than the healthy population. This result is consistent with reports of reduced head movement during depression [19]. From the basic emotions, average sadness intensity was the only one that demonstrated the same trend between males and females, with both genders expressing significantly less sadness while suffering from depression or PTSD.

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 Fig. 3a where the gender has an effect on the Emotional Variance indicator. Females over all, in both distressed and non-distressed conditions seem to showcase more emotional variability than males, a result that agrees with literature and previous observations in clinical settings [13]. All these observations serve as a good indication that a gender- dependent approach will benefit the assessment of depression and PTSD.

6 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 Fig. 4 we demonstrate the basis of our hypothesis on two indicators. One can see that the indicator of average disgust intensity loses effect when studied in a gender-independent approach. In the following sub-sections we describe the compared models, the methodology we follow for the classification experiment and present our results.

6.1 Models

In the experiments we evaluate the performance of 3 models: **Baseline** which uses the majority vote where all observations

Table 3 Example of behavior indicators showing gender similarities in trend

	Feature	Gender	Hedge's g	*p* value	trend
Depression	HeadRot Var	Males	−0.57	0.185	↓
		Females	−0.89	**0.047**	↓
	Emotion Var	Males	−0.59	0.166	↓
		Females	−0.67	0.128	↓
PTSD	HeadRot Var	Males	−0.74	0.054	↓
		Females	−0.59	0.175	↓
	Emotion Var	Males	−0.39	0.299	↓
		Females	−0.89	**0.045**	↓

Bold values indicate $p \leq 0.05$

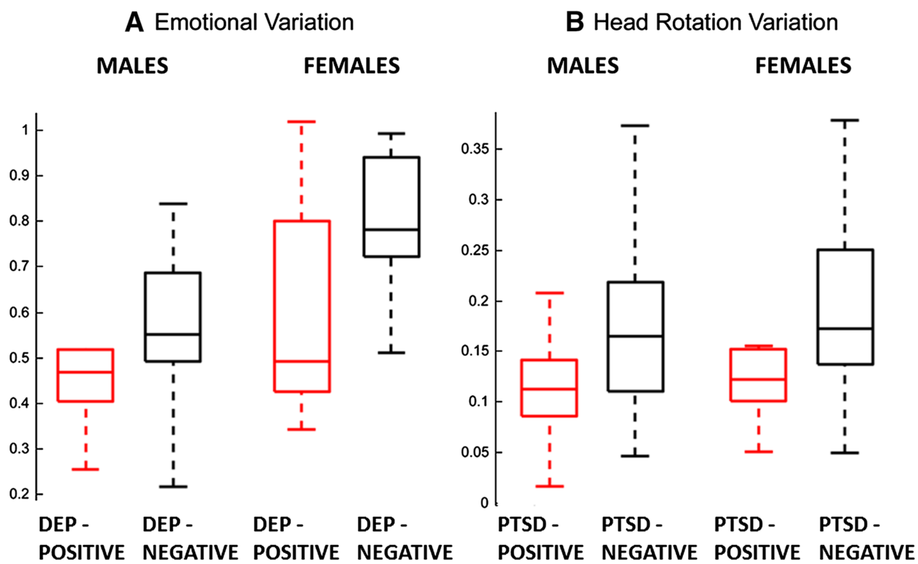

Fig. 3 Example of behavior indicators showing gender similarities in trend. In both the cases of **a** Emotional Variation in depression and **b** Head Rotation Variation in PTSD, the conditions show same trends among genders.

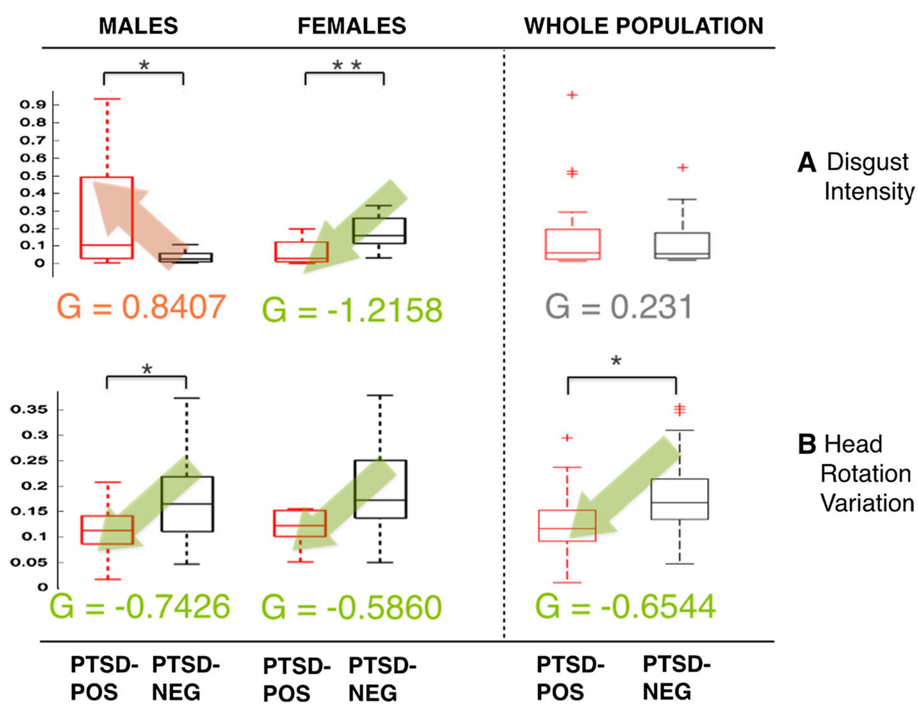

Fig. 4 Demonstrating the interaction of gender and a psychological condition (PTSD) and how this affects studying the whole population. In case (**a**) of measured disgust intensity we observe a cross-over interaction of gender and the psychological condition of PTSD. One can see that as a result the effect of PTSD on that indicator is lost when one looks at the population as a whole. In case (**b**) however, where we measure the average head rotation variation, the effect of PTSD on that indicator is similar between the two genders and thus is preserved when we look at the population as a whole

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 'Males' and 'Females'. 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.

6.2 Feature representation

Using the automatic sensing framework described in Sect. 4.C we extracted the behavioral signals and computed basic summary statistics for each interaction in our dataset (Sect. 3). 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 AUs 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 emo-

Table 4 Classification results for Depression and PTSD. We show that a gender-dependent model performs better than a gender-independent one

	Population	Baseline F1	Gender-independent F1	Gender-dependent F1
Depression	All	0.576	0.722	–
	Males	0.610	0.756	**0.808**
	Females	0.512	0.682	**0.858**
PTSD	All	0.540	0.785	–
	Males	0.579	0.739	**0.811**
	Females	0.533	0.840	**0.908**

Bold values indicate $p \leq 0.05$

tions mentioned in Sect. 4.2. 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 AUs (AU4-AU7-AU9). Our final feature pool contained 20 features.

6.3 Classification

As a simple approach, we chose a Naive Bayes classifier [2] which has the advantage of having a limited number of hyper-parameters. 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 versus non-PTSD or depressed versus 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).

6.4 Results

In Table 4 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 'Males' and 'Females'. Also, the gender-independent approach performs better than the baseline for all test groups.

6.4.1 Selected feature analysis

In PTSD classification, the affect of anger, contempt and some combination of the AU4-AU7-AU9 was prominent. Also the emotional variation was selected in both the gender models for gender- dependent classification. In depression neutral face measure and frowning played important role. In general the gender- independent approach selected fewer features than the gender- dependent models and interesting was the case of gender- independent depression classifier which seemed to perform the best for the whole population using only one feature: the neutral intensity which measures flatness of expression.

To showcase the benefit of a gender-dependent approach, we look at the PTSD-classifier margins for one of these indicators, AU4, when used in the gender-independent model: 0.0985, and when used in the gender-dependent approach: -0.7326 and 0.6070 for men and women respectively. It seems that this indicator can offer discriminative information for PTSD when we take the gender- dependent approach, but not as much in a gender-independent one.

7 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 [13]. There are intrinsic differences in nonverbal behaviors among genders [36], sometimes amplified or attenuated by social norms and gender-related expectations [37,38]. However, one should be cautious about the interpretation of such phenomena. For one, elicited behaviors are often influenced by the interaction style [39] 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. Besides the different challenges introduced by sex (ex. facial hair on men),

[2] http://www.mathworks.com/products/statistics.

shape and appearance features may be influenced by gender, so one must be careful when constructing features that rely on those. In the feature list we chose for our analysis we chose to represent information that should not rely on gender like head movement, and AUs that are globally defined.

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 [40]. As mentioned in Sect. 3.2 by introducing a virtual human as the interviewer we are attempting to standardize the interaction and minimize the variability of the given stimuli. For example, the virtual human can deliver the same question in the exact same way (executed with the same controlled animation and verbalized in the exact same way) among participants. This way one can reduce the effect of one more external factor in the dyadic interaction and hopefully isolate differences in behavior that are caused by other factors such as psychological illness or other inherent motivations. During our experiments, we used consistently the same female virtual human (Ellie). Understandably, this choice can have a different effect on the different genders, but at this point we prioritized having a consistent setup.

In addition to the above gender related differences, some psychological conditions like depression and PTSD have different base rates among the two genders [41], 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. We would like to point out that besides increasing the PTSD or depression classification accuracy on the whole population, we are hoping that the gender analysis can help shed some more light into the internal motivations that cause a distressed person to express or avoid certain behaviors. One interesting result, showcasing less sadness while in depression seems non-intuitive but fits very well the nonverbal social withdrawal hypothesis as seen in previous work [19]. In general, we observed most gender differences in the category of expressions of emotions, while head movement variation showed similarities. Perhaps motivations to express or avoid certain communicative behaviors are subject to different mental processes in men and women while in depression, whereas reduction of overall movement (i.e. motor retardation) is a neurological effect of the illness common in both genders. The work presented in this article does not attempt to answer such questions, only to motivate them by gender-dependent analysis.

At this point, we would like to mention that the introduced gender-dependent approach does not hinter nor discourages 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 [34]). By using the first 3sec of the video 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 [10]. 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 [42]. 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.

8 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-ependent 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.

One possible future direction of this work is to analyze the context of interaction and the effect of affective stimuli on the discriminative power of behavioral indicators. Further, we plan to explore indicators based on a dynamic and multimodal observations by incorporating additional modalities, such as audio, body gestures/posture as well as context/lexical patterns.

Acknowledgments 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. Baltrusaitis T, Robinson P, Morency L (2012) 3d constrained local model for rigid and non-rigid facial tracking. In: CVPR, 2012 IEEE Conference
2. Littlewort G, Whitehill J, Wu T, Fasel I, Frank M, Movellan J, Bartlett M (2011) The computer expression recognition toolbox (cert). In: 2011 IEEE International Conference on FG 2011
3. Ekman P, Matsumoto D, Friesen WV (1997) Facial expression in affective disorders. What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS) 2:331–342
4. Ellgring H (1989) Nonverbal communication in depression. Cambridge University Press, Cambridge
5. Cohn JF, Kruez TS, Matthews I, Yang Y, Nguyen MH, Padilla MT, Zhou F, De la Torre F (2009) Detecting depression from facial actions and vocal prosody. In: 3rd international conference on ACII 2009, pp 1–7
6. McIntyre G, Gocke R, Hyett M, Green M, Breakspear M (2009) An approach for automatically measuring facial activity in depressed subjects. In: 3rd international conference on ACII 2009
7. Maddage MC, Senaratne R, Low L-SA, Lech M, Allen N (2009) Video-based detection of the clinical depression in adolescents. In: EMBC 2009. Annual International Conference of the IEEE, pp 3723–3726
8. Diagnostic and Statistical Manual of Mental Disorders DSM-IV-TR Fourth Edition (Text Revision). Amer Psychiatric Pub, 4th edition, July 2000
9. Diagnostic and Statistical Manual of Mental Disorders DSM-V Fifth Edition. American Psychiatric Association, 5th edition, June 2013
10. Russell JA, Barrett LF (1999) Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant. J Personal Soc Psychol 76(5):805–819
11. Ian Reed L, Sayette MA, Cohn JF (2007) Impact of depression on response to comedy: a dynamic facial coding analysis. J Abnorm Psychol 116:804–809
12. Kirsch A, Brunnhuber S (2007) Facial expression and experience of emotions in psychodynamic interviews with patients with ptsd in comparison to healthy subjects. Psychopathology 40(5):296–302
13. Troisi A, Moles A (1999) Gender differences in depression: an ethological study of nonverbal behavior during interviews. J Psychiat Res 33(3):243–250
14. Wolfsdorf BA, Mattia JI, Zlotnick C, Zimmerman M (2001) Gender differences in patients with posttraumatic stress disorder in a general psychiatric practice. Am J Psychiatry 158(11):1923–1925
15. Draijer N, Gersons BP, Olff M, Langeland W (2007) Gender differences in posttraumatic stress disorder. Psychol Bull 133(2):183–204
16. Green B (2003) Post-traumatic stress disorder: symptom profiles in men and women. Curr Med Res Opin 19(3):200–204
17. Cummins N, Joshi J, Dhall A, Sethu V, Goecke R, Epps J (2013) Diagnosis of depression by behavioural signals: a multimodal approach. In: Proceedings of the 3rd ACM international workshop on audio/visual emotion challenge, AVEC '13, pp 11–20, New York, NY, USA, 2013. ACM
18. Meng H, Huang D, Wang H, Yang H, AI-Shuraifi M, Wang Y (2013) Depression recognition based on dynamic facial and vocal expression features using partial least square regression. In: Proceedings of the 3rd ACM International Workshop on Audio/Visual Emotion Challenge, AVEC '13, pp 21–30, New York, NY, USA, 2013. ACM
19. Girard JM, Cohn JF, Mahoor MH, Mavadati SM, Hammal Z, Rosenwald DP Nonverbal social withdrawal in depression: evidence from manual and automatic analyses. Image and Vision Computing
20. D'Mello S, Graesser A (2012) Dynamics of affective states during complex learning. Learn Instr 22(2):145–157
21. Joshi J, Dhall A, Goecke R, Cohn JF (2013) Relative body parts movement for automatic depression analysis. In: Proceedings of the 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII '13, pp 492–497, Washington, DC, USA, 2013. IEEE Computer Society
22. Scherer S, Stratou G, Mahmoud M, Boberg J, Gratch J, Rizzo A, Morency L-P (2013) Automatic behavior descriptors for psychological disorder analysis. In: IEEE conference on automatic face and gesture recognition, 2013
23. Blanchard EB, Jones-Alexander J, Buckley TC, Forneris CA (1996) Psychometric properties of the ptsd checklist (pcl). Behav Res Therapy 34(8):669–673
24. Kroenke K, Spitzer RL (2002) The phq-9: a new depression and diagnostic severity measure. Psychiatr Ann 32:509–521
25. Kroenke K, Spitzer RL, Williams JBW (2001) The phq-9. J Gen Intern Med 16(9):606–613
26. Bolton EE, Gray MJ, Litz BT (2006) A cross-lagged analysis of the relationship between symptoms of ptsd and retrospective reports of exposure. J Anxiety Disord 20(7):877–895
27. Hoge CW, Castro CA, Messer SC, McGurk D, Cotting DI, Koffman RL (2004) Combat duty in iraq and afghanistan, mental health problems, and barriers to care. N Engl J Med 351(1):13–22
28. Andrykowski MA, Cordova MJ, Studts JL, Miller TW (1998) Posttraumatic stress disorder after treatment for breast cancer: prevalence of diagnosis and use of the ptsd checklistcivilian version (pclc) as a screening instrument. J Consult Clin Psychol 66(3):586–590
29. Stefan S, Giota S, Jonathan G, Morency L-P (2013) Investigating voice quality as a speaker-independent indicator of depression and ptsd. In: Proceedings of interspeech
30. DeVault D, Georgila K, Artstein R, Morbini F, Traum D, Scherer S, Rizzo A, Morency L-P (2013) Verbal indicators of psychological distress in interactive dialogue with a virtual human. In SIGDIAL, Metz, France, August
31. Ekman P, Rosenberg EL (2005) What the face reveals: basic and applied studies of spontaneous expression using the facial action coding system (FACS) (series in affective Science). Oxford UP, USA, New York
32. Bonanno GA, Keltner D (1997) Facial expressions of emotion and the course of conjugal bereavement. J Abnorm Psychol 106:126–137
33. Morency L, Whitehill J, Movellan J (2008) Generalized adaptive view-based appearance model: Integrated framework for monocular head pose estimation. In: 8th IEEE International Conference on FG '08
34. Kublbeck C, Ernst A (2006) Face detection and tracking in video sequences using the modifiedcensus transformation. Image Vis Comput 24(6):564–572
35. Hedges LV (1981) Distribution theory for glass's estimator of effect size and related estimators. J Educ Behav Stat 6(2):107–128
36. Deutsch FM, LeBaron D, Fryer MM (1987) What is in a smile? Psychol Women Q 11(3):341–352
37. Hess U, Adams RB, Kleck RE Jr (2004) Facial appearance, gender, and emotion expression. Emotion 4(4):378
38. Adams R Jr, Kleck R, Hess U (2005) Who may frown and who should smile? dominance, affiliation, and the display of happiness and anger
39. Krämer NC, Hoffmann L, Kopp S (2010) Know your users! empirical results for tailoring an agent's nonverbal behavior to different user groups. In: Proceedings of IVA, IVA'10, pp 468–474

40. Kulms P, Kramer NC, Gratch J, Kang S-H (2011) It's in their eyes: a study on female and male virtual humans gaze. In: IVA, volume 6895 of, Lecture Notes in Computer Science, pp 80–92
41. Rief W, Nanke A, Klaiberg A, Braehler E (2004) Base rates for panic and depression according to the brief patient health questionnaire: a population-based study. J Affect Disord 82(2):271–276
42. Campbell DG, Felker BL, Liu C-F, Yano EM, Kirchner JE, Chan D, Rubenstein LV, Chaney EF (2007) Prevalence of depressionptsd comorbidity: implications for clinical practice guidelines and primary care-based interventions. J Gen Intern Med 22(6):711–718