Investigating the Interplay Between Self-Reported and Bio-Behavioral Measures of Stress: A Pilot Study of Civilian Job Interviews with Military Veterans

1st Ehsanul Haque Nirjhar
Computer Science and Engineering
Texas A&M University
College Station, TX, USA
nirjhar71@tamu.edu

2nd Md Nazmus Sakib
Multidisciplinary Engineering
Texas A&M University
College Station, TX, USA
mnsakib@tamu.edu

3rd Ellen Hagen
Psychological & Brain Sciences
Texas A&M University
College Station, TX, USA
elli_hagen2346@tamu.edu

4th Neha Rani
Computer Science
University of Florida
Gainesville, FL, USA
neharani@ufl.edu

5th Sharon Lynn Chu
Computer Science
University of Florida
Gainesville, FL, USA
slchu@ufl.edu

6th Winfred Arthur, Jr.
Psychological & Brain Sciences
Texas A&M University
College Station, TX, USA
w-arthur@tamu.edu

7th Amir H. Behzadan
Construction Science
Texas A&M University
College Station, TX, USA
abehzadan@tamu.edu

8th Theodora Chaspari
Computer Science and Engineering
Texas A&M University
College Station, TX, USA
chaspari@tamu.edu

***Abstract*—Transitioning from the military to the civilian lifestyle, especially for military veterans who decide to pursue careers in the civilian workforce, is often a difficult experience. The job interview, a task in which the interviewees meet and discuss their skills and career goals with strangers in a position of authority, is the first step of assimilation into the civilian workplace, which might cause them to experience nervousness or anxiety. This feeling of excessive stress may compromise the interviewee's performance, therefore potentially impeding their successful transition to the workforce. Intelligent interview training technologies would benefit from automated stress detection systems that could assist interviewees in better understanding causes and antecedents of stressors during their interaction with the interviewer. This paper examines self-reported and bio-behavioral measures of stress experienced during mock job interviews conducted with 24 U.S. military veterans. Self-reported measures were captured via a global measure of stress reported by the participant at the conclusion of the interview, and a continuous moment-to-moment annotation of stress resulting from the retrospective inspection of the interview video recording. Bio-behavioral indices of stress include physiological reactivity measures captured via electrodermal activity and electrocardiogram signals, as well as acoustic measures extracted from speech. Results indicate that physiological reactivity measures exhibit moderate-to-strong correlation with self-reported measures of stress, and can be thus used to estimate the self-reported stress measures. Augmenting the feature space with demographic and psychological traits can further improve the accurate detection of stress during the interviews.**

***Index Terms*—workforce reskilling, job interview, U.S. veterans, stress, physiological signals, speech, individual differences**

The authors would like to acknowledge the National Science Foundation (NSF) for funding this work (NSF #1956021, NSF #1956087).

I. Introduction

Currently, there are more than 18 million veterans of the U.S. Armed Forces, whose previous service experience ranges from World War II to the more recent combat missions in Iraq and Afghanistan. This accounts for about 7% of the U.S. adult population [1], [2]. The transition to civilian life is described as difficult by over two-thirds of the veteran population [3], [4], and several studies have pointed to this step as daunting for many U.S. veterans primarily as a result of the personal and cultural adjustments expected during the process [5], [6]. There are multiple facets of transitioning into the civilian life, including ensuring physical and mental well-being, dealing with the trauma, reconnecting with friends and family, accessing proper support systems, and finding reasonable employment opportunities. Despite having a very strong work ethic, dedication, and experience, 6.5% of the U.S. veterans were unemployed in 2020, which was up from 2.9% in the year before [7], a number that potentially reflects the impact of the COVID-19 pandemic on the job market. In addition, finding a civilian job has been reported as one of the major challenges faced by veterans. In a survey conducted by Prudential Financial [4], 69% of the respondents stated that finding a career in the civilian sector is the greatest challenge they had encountered while transitioning. Some of the factors that could contribute to this problem include the negative stereotype toward veterans [8]–[10], as well as the communication gap and cultural differences between the military and civilian workplace [11]–[13]. These barriers are faced by veterans in different phases of the civilian employment, as early as the interview stage, which serves as the first hurdle in job search.

978-1-6654-5908-2/22/$31.00 ©2022 IEEE

A large body of literature exists on U.S. veterans' overall experience when transitioning from the military to civilian job sector [11], [14]. This prior work mainly focuses on the difficulties faced by the veterans due to the cultural difference between the military and civilian institutions, and the comparative lack of defined structure in civilian jobs. The interview is a crucial step in obtaining a job, and therefore, an important factor in the transition of veterans. Hence, interviews may elicit stress among the veterans, which can impede their successful transition [15]–[17]. Assistive technologies (i.e., intervention tools, interview training modules) can help veterans overcome this obstacle. The automated analysis of job interviews through multimodal features (e.g., speech, physiology, language, video) is the first step in developing such technologies. Prior work in the field of automated job interview analysis has been mostly conducted using data from college students and concentrates on inferring the hirability and performance metrics of the interviewees [18]–[20]. Very few papers have explored the extent to which the detection of stress can be performed through automated interview analysis [17], [21]. In addition, research in affective computing has studied algorithms to automatically detect stress based on various bio-behavioral indices. However, stress in these studies is mostly elicited via theoretically grounded stressors, such as a mental arithmetic task, stressful event recall, public speaking, or driving [22]–[24]. The job interview encounter has not been commonly used as a stress elicitation method in general, let alone its impact on a relatively unexplored population, such as military veterans.

In this paper, we present a pilot study that aims to bridge this gap. We collected data from 24 U.S. veteran participants, who participated in a mock job interview with civilian sector hiring personnel. Bio-behavioral signals (i.e., electrodermal activity (EDA), electrocardiogram (ECG), speech) were obtained during and prior to the interview through wearable sensors. In addition, individual difference pertaining to demography (e.g., age, sex, duration of military service) and psychological characteristics (e.g., personality traits, trait anxiety) were also obtained using a set of self-report measures. Participants self-reported their stress levels via a global score obtained at the end of the interview and a moment-to-moment annotation that was obtained after retrospectively inspecting the video recording of their interview. We examine the association between the stress measures and the bio-behavioral indices, and aim to find factors contributing to stress from the pool of demographics and psychological traits. In summary, this paper addresses the following research questions:

RQ1: To what extent does the considered mock interview induce stressful responses and are these manifested in bio-behavioral measures such as physiology and speech?

RQ2: Are global and moment-to-moment self-reported stress measures correlated, and how are these self-reported measures associated with bio-behavioral measures of stress and individual difference measures?

RQ3: Can the self-reported stress be estimated by the combination of bio-behavioral and individual difference measures?

II. Related Work

There is a significant difference in culture between civilian and military life and work [5], [25]. This makes the transition from military to civilian life very challenging in many aspects. A study on veterans returning from Afghanistan and Iraq reported that veterans felt a disconnection with the civilian life while transitioning [5]. In a study of Gulf War II veterans, 64% of respondents described their transition as a difficult experience, and almost half did not feel prepared for it [4]. Among the different stages of this transition, finding a job was reported as the greatest challenge by 69% of respondents. Another major challenge faced by veterans is translating their learned skills in the military to the civilian workspace. Veterans gain a wide array of experience during their military service [8]. In a study conducted by Prudential Financial [4], 60% of veterans responded to the question "how their military experience translates to skills of interest to a civilian employer" by citing it as the biggest challenge in finding civilian employment. This is also evident from the findings of other studies [9], [11]. The communication process is very different in these two workspaces [12]. This may result in a communication gap between veterans and their civilian peers, even during the job interview, which creates a barrier for the transitioning veterans. All these challenges may cause veterans to feel additional stress during their civilian job interviews, as interviews tend to elicit stress [15], [17]. Despite these unique challenges, the way stress is experienced and perceived by this population during the civilian job interview has not been extensively examined before.

Prior work on job interview analysis is mostly centered around the computational inference of interviewee hirability and performance. Nguyen *et al.* explored the use of non-verbal behavioral indices to estimate several hirability metrics (e.g., hiring decision, communication skills, stress management) [19]. They compared their model's performance with psychometric measure responses and found that hirability is better estimated with bio-behavioral indices (e.g., prosodic cues, optical flow measures) that can better capture the interaction during the interview. Chen *et al.* further augmented prosodic and facial expression features with interviewees' linguistic features to quantify an interviewee's personality traits and hirability score [20]. In their TARDIS framework for social coaching for job interviews, Anderson *et al.* trained a Bayesian Network using manually annotated non-verbal social cues to predict whether the interviewees are feeling stressed [21]. Naim *et al.* further attempted to estimate various social traits and interview-related behaviors (i.e., stress, engagement, structured answer) along with hirability and performance ratings using job interview videos. They observed that estimation of stress yielded low accuracy due to unreliable ground truth from crowd-sourced ratings. Moreover, findings suggested that a controlled and realistic experiment setting might be required to elicit adequate levels of stress in the interviewees who were college students. Although not prevalent in job interview analysis, the use of physiological signals is very common in the stress detection literature. EDA

TABLE I
SUMMARY OF THE DEMOGRAPHIC INFORMATION OF THE PARTICIPANTS.

Demographic Factors	Value ($N = 24$)
Age (yrs)	36.4 (10.6)
#Female participants	2
Range of #deployments	0–8
Duration of military service (yrs)	9.0 (5.9)
Duration since separated from the service (yrs)	8.9 (9.0)
#Participants attending transition assistance class	8

*X(Y) = mean (SD)

and ECG signals have been used to detect stress in various stressful situations, such as driving [23], lab studies [24], public speaking [26], and work settings [27]. However, the use of physiological measures for detecting stress during job interviews has not been adequately explored.

This paper aims to address gaps in better understanding stress mechanisms during the job interview via a pilot study conducted with 24 U.S. veterans. The main contributions of this work are: (1) We explore the extent to which stress can be elicited via a mock job interview setting; (2) Beyond the typically employed self-reports administered before and after the stressor tasks, we further quantify stress in a continuous manner based on interviewees' retrospective inspection of the interview video; and (3) We investigate whether stress can be estimated using bio-behavioral measures of physiology and speech, and the extent to which it depends on the demographic and psychological traits of this unique population.

III. DATA

Our data comes from an ongoing user study involving mock job interviews. This study has two sets of participants– Interviewees, who are U.S. veterans transitioning (or transitioned) to the civilian life ($n = 24$), and Interviewers, who are industry representatives with prior experience in conducting interviews and recruiting personnel ($n = 5$). The study has been reviewed and approved by the Institutional Review Board (IRB) of Texas A&M University (IRB #2020-0709D). U.S. veteran participants were recruited via campus-wide emails, and advertisement, and the industry representative participants were recruited through professional contacts of the team. The study was conducted in a hybrid format. Interviewees were present in the lab, while interviewers joined the session remotely. Interviews were conducted via Zoom video conferencing [28]. In this paper, we discuss the data from U.S. veterans only, who will be referred to as participants in subsequent text. The demographic information of the 24 participants who have so far completed the study is summarized in Table I. Prior to the interview session, participants shared their résumé with the research team. Based on the information contained in each résumé, a customized mock job posting was created for each participant. Participants were asked to consider that they applied for the given job posting and they were going for an interview for that job. Both the résumé and the mock job posting were shared with the interviewers, who were asked to conduct the interviews accordingly as they would normally do as part of their work. These steps were taken to ensure a naturalistic conversation during the mock job interviews.

During the day of the interview, participants wore two wearable devices that captured their physiological signals– the Empatica E4 wristband [29] and the Actiheart 5 chest-worn device [30]. The E4 wristband recorded EDA sampled at 4 Hz. The Actiheart 5 is a single-lead ECG recording device, that collects ECG data at 512 Hz. At the beginning, participants completed the demography questionnaires (e.g., age, biological sex, year in military service, last year of service), and prior daily experience (e.g., meal, caffeine/alcohol intake). Next, they completed a set of measures which were used to operationalize the psychological traits. These consisted of an IPIP measure of the Five factor model of personality (IPIP) [31], Trait scale of State-Trait Anxiety Inventory (STAI Trait) [32], Coping Inventory for Stressful Situations-Short Form (CISS-21) [33], Symptom Checklist 90-Revised (SCL-90-R) [34], and Life Event Checklist for DSM IV (LEC) [35]. The IPIP measures the five factor model of personality traits. The STAI Trait captures the trait-based anxiety, while CISS-21 captures what strategies individuals use during different stressful situations. Next, SCL-90-R records various psychological symptoms and distress, and LEC screens for potentially traumatic events in an individual's life. Finally, a general mental ability (GMA) test [36] was administered to measure cognitive ability. All these measures quantify the wide range of psychological traits among the participants.

After completing the questionnaires, participants completed a relaxation session in which they watched a video of natural images with soothing music for 10 minutes to obtain their physiological reactivity at rest. Next, they completed the State scale of State-Trait Anxiety Inventory (PRE STAI State) [32] that captured their anxiety as a current state just before the interview session. Next, participants were introduced to the interviewer, who was connected through Zoom. The interview session was recorded. The average duration of the interviews was approximately 18 minutes (SD = 6.35 minutes). After the interview, participants completed another set of measures that included the State scale of State-Trait Anxiety Inventory (POST STAI State) [32] to capture their immediate anxiety state after interview, and Interview Experience Survey to record their thoughts about their performance in the interview. Finally, participants watched their recorded interview and performed moment-to-moment continuous rating of their stress level during the interview on a scale from 1 (No Stress) to 5 (High Stress) using the CARMA annotation tool [37].

IV. METHODOLOGY

A. Data Pre-processing

Physiological signals collected during the study are susceptible to noise, due to sensor displacement, motion artifacts, and loss of contact. Therefore, pre-processing is necessary for noise suppression and artifact removal. An initial visual inspection of the raw EDA and ECG signals is performed to identify the signals that do not have the expected shape (i.e., phasic and tonic responses in EDA [38], presence of QRS complex in ECG [39]). EDA signals that continuously display no fluctuation and very low value (i.e., $< 0.01 \mu S$) are discarded. Similarly, ECG signals with flat-line shapes are also removed from the data. This resulted in exclusion on 5 and 7 participants' ECG data from Interview and Relaxation sessions, respectively.

After the visual inspection, outliers are removed from the EDA signal and are replaced via linear interpolation. A window size of 48 samples (i.e., 12 seconds) is used for this purpose. Next, high frequency noise is removed using a Bateman low-pass filter of 2-second window length. Meanwhile, ECG signals are filtered using a finite-impulse response bandpass filter, followed by R-peak detection using the BioSPPy toolbox [40]. Next, outliers caused by the motion artifact and sensor displacement are automatically replaced from the sequence of R-R intervals, followed by the removal of ectopic beats using hrv-analysis toolbox [41]. Finally, EDA and ECG signals are segmented into relaxation and interview phases based on the start and end timestamps of the sessions. Audio signals from the zoom recording contain transcripts with timestamps for all speakers. Voice activity detection (VAD) is performed at the timestamps where only the participants are speaking, using OpenSMILE [42]. These segments are further used for acoustic feature extraction.

B. Feature Extraction

A total of 4 EDA features are extracted from the preprocessed EDA signals using Ledalab [43]. These features include the mean and standard deviation of skin conductance level (SCL), and the skin conductance response (SCR) amplitude and frequency. The EDA signal is controlled by sympathetic nervous system (SNS), which provides the "fight-or-flight" response during times of distress [38]. Thus, SCL and SCR features tend to display increased reactivity during stress [22]. Additional features are extracted from the R-R interval series obtained from ECG signals using hrv-analysis toolbox [41]. ECG features contain 16 time-domain and 7 frequency-domain features. Time-domain features include mean, standard deviation, median, and range of the N-N interval (NNI), mean, standard deviation, minimum and maximum heart rate (HR), number (and proportion) of interval differences of successive R-R intervals greater than 20 and 50 ms (NNI20, PNNI20, NNI50, PNNI50), and square root of the mean of the sum of the squares of differences between adjacent N-N-intervals (RMSSD). Frequency-domain features consist of the very low frequency (VLF), low frequency (LF), and high frequency (HF) power components, LF-HF ratio, and total power. NNI, HR, and LF features mostly reflect the activity of the SNS, while RMSSD, HF, and LF-HF mostly reflect the activity of the parasympathetic nervous system (PNS) contributing towards self-regulation after experiencing distress [44], [45]. Therefore, these features are vital indicators in identifying stress occurrence and regulation. Finally, OpenSMILE is used to extract acoustic features from the participants' audio signal [42]. We extracted the extended Geneva Minimalistic Acoustic Parameter Set (eGeMAPS), as this feature set is concise in size and is widely used in the affective computing domain. The dimension of this feature set is 88. These features were computed over a 30-millisecond window and then averaged over the speech segments of the participants during the interview.

C. Measures of Stress

Participants self-reported their stress before and after the interview. These scores are referred to as PRE STAI State and POST STAI State, respectively, which comprise the global stress scores. We also calculated the difference in STAI State scores after the interview. In addition, participants completed a moment-to-moment annotation by watching their interview recording. To compare them with self-reported stress measures, a set of metrics are calculated from their annotation time series. These include the mean of the rating (Rmean), mean of the the peak values (Rpeak), and frequency of peaks (PeakFreq), computed over the entire interview session. These metrics are considered as moment-to-moment measures of stress. Association between various measures of stress and the bio-behavioral measures are further explored.

D. Individual Difference Measures

Individual differences were operationalized using the measures mentioned in Section III. We used 6 demographic measures for further experimentation, namely age, biological sex, number of deployments, whether participant attended transition assistance class, duration of service, and number of years since the participant had separated from service. These measures include general and veteran-specific information. In addition, 7 psychological traits were obtained from the IPIP, STAI Trait, and SCL-90-R, specifically, extraversion, agreeableness, conscientiousness, emotional stability, and openness scores from IPIP, general trait anxiety from STAI Trait, and Anxiety score from SCL-90-R. We examine the extent to which these measures moderate the association between stress measures and bio-behavioral indices.

E. Evaluation

In order to answer RQ1, a paired t-test is conducted to check whether the various measures of stress depict significant differences before and after the interview. This can contribute to determining the efficacy of the mock interview in eliciting stress. PRE and POST STAI State scores, and features obtained from physiological signals (i.e., EDA, ECG) during relaxation and interview sessions are used for the t-test. Speech features and moment-to-moment stress score metrics are available only for the interview session. Therefore, they are not included in this study.

To respond to RQ2, we examine the association between different stress scores (i.e., POST STAI State, Difference in STAI State, Rmean, Rpeak, PeakFreq) (Section IV-C). Next, we compute the correlation between the various bio-behavioral indices computed from the EDA, ECG, and Speech signals (Section IV-B) and the stress scores (Section IV-C). Moreover, we fit a linear regression model with stress scores as the dependent variable and the combination of demographic and psychological trait measures as independent variables. By inspecting the weights obtained from this regression, we discuss how self-reported stress varies with these individual difference measures.

Finally, for RQ3, we use a linear regression model to estimate the self-reported stress based on different combinations of input variables. These include each feature group alone (i.e., EDA, ECG, Speech), as well as their combination with the individual difference measures. Feature dimensions of EDA, ECG, speech, demographic, and psychological traits are 4, 23, 88, 6, and 7, respectively. This experimentation is conducted in a leave-one-subject-out manner. The Pearson's correlation coefficient (r) between the actual and estimated stress scores is used as an evaluation metric.

V. Results

We now discuss the results of the experiments presented alongside each of the research questions (Section I).

A. Effectiveness of Mock Job Interviews in Eliciting Stress

Table II presents the results of the paired t-test with respect to the various measures of stress before and after the interview. Results indicate that although the POST STAI State score (mean = 30.58, SD = 8.66) has a higher average than PRE STAI State score (mean = 28.96, SD = 7.43), the difference between the two is not significant ($p = 0.34$). Meanwhile, significant differences can be observed with respect to the EDA features captured during relaxation and interview sessions. Mean SCL during the relaxation session (mean = 1.99 μS, SD = 2.70 μS) is significantly lower compared to mean SCL during the interview (mean = 3.62 μS, SD = 4.15 μS). The SCR frequency also depicts a higher value in the interview (mean = 10.65 peaks/minute, SD = 5.57 peaks/minute) than the relaxation session (mean = 4.08 peaks/minute, SD = 3.36 peaks/minute). The obtained result is in line with the literature, since the EDA features are expected to increase during a stressor stimulus [22], [38]. Among the ECG features, time-domain features like mean NNI and mean HR display a significant difference between the relaxation and interview session, while the frequency domain features do not show any such difference. Mean NNI in relaxation (mean = 836.48 ms, SD = 180.22 ms) is significantly higher than in interview (mean = 785.54 ms, SD = 136.94 ms), while mean HR shows the significant increase from relaxation (mean = 75.72 bpm, SD = 16.09 bpm) to interview session (mean = 79.58 bpm, SD = 14.47 bpm). HR and NNI depict an inverse relationship, and HR tends to rise with stress [22], which is evident in the result. Results of this experiment suggest that the mock interview session in the pilot study is capable of inducing stress and it is manifested in the bio-behavioral indices.

B. Correlation between Self-reported and Bio-behavioral Stress Measures

In response to RQ2, the Pearson's correlation coefficient (r) is calculated between the various stress measures. First, the association between global stress measures (i.e., STAI State) and moment-to-moment stress measures (i.e., Rmean, Rpeak, PeakFreq) has been examined to explore the internal consistency of the measures, as both are used for retrospectively rating stress. Results are shown in Table III. Almost all the moment-to-moment stress measures exhibit moderate-to-strong correlation with the global stress measures– POST STAI State score and the difference in STAI State scores. This indicates that the continuous stress annotation, obtained via watching the interview videos, is consistent with the global stress measures captured through the STAI State scores.

Next, we examine the correlation between bio-behavioral measures of stress with the global and moment-to-moment measures of stress. Results in Table IV indicate that all the EDA features are positively correlated with POST STAI State score (e.g., $r = 0.38$, $p = 0.07$ for mean SCL). To the contrary, only the SCR frequency exhibits a moderate correlation with the moment-to-moment stress measures (e.g., $r = 0.36$, $p = 0.08$ with Rmean). Next, among the ECG features, Max HR exhibits a strong positive correlation with moment-to-moment stress measures (i.e., $r = 0.69$, $p < 0.01$ with Rpeak), while mean NNI presents a strong negative correlation (i.e., $r = -0.50$, $p < 0.05$ with Rpeak). The Max HR measure also depicts a strong correlation with the POST STAI State score (i.e., $r = 0.45$, $p = 0.05$ with Rpeak). These findings are consistent with prior work [22], [26]. The RMSSD and LF-HF ratio measures display a significantly moderate correlation with the POST STAI State score. However, the direction of these correlations is opposite to the expected one [22], potentially due to the fact that these measures might require longer analysis windows [46] to yield reliable estimates of the parasympathetic activity. Finally, speech features mostly exhibit zero-to-very low correlations, except the spectral slope and α ratio features showing weak positive correlations with the stress measures. Such low correlations may result from the aggregation of speech features over the whole interview session. Speech features tend to change rapidly, therefore aggregating them over large window may fail to capture their temporal variation due to stressors. Finally, we examine the effect of individual factors on state anxiety. For this purpose,

TABLE II
T-TEST RESULTS IDENTIFYING SIGNIFICANT DIFFERENCES BETWEEN RELAXATION AND INTERVIEW SESSIONS WITH RESPECT TO BIO-BEHAVIORAL AND SELF-REPORTED MEASURES.

Modality	Measure	t-test results
Self-report	STAI State ($N = 24$)	$t(23) = -0.97$
EDA	Mean SCL ($N = 24$)	$t(23) = -3.12^{**}$
	SD SCL ($N = 24$)	$t(23) = -3.09^{**}$
	SCR Amplitude ($N = 24$)	$t(23) = -1.76$
	SCR Frequency ($N = 24$)	$t(23) = -7.76^{**}$
ECG	Mean NNI ($N = 17$)	$t(16) = 3.01^{**}$
	SDNN ($N = 17$)	$t(16) = 0.37$
	NNI50 ($N = 17$)	$t(16) = -2.81^{*}$
	Mean HR ($N = 17$)	$t(16) = -3.21^{**}$
	LF ($N = 17$)	$t(16) = 0.49$
	HF ($N = 17$)	$t(16) = -0.24$
	LF-HF ratio ($N = 17$)	$t(16) = -0.09$

*: p <0.05, **: p <0.01

TABLE III
PEARSON'S CORRELATION COEFFICIENT BETWEEN SELF-REPORTED GLOBAL STRESS MEASURES AND MOMENT-TO-MOMENT STRESS MEASURES.

Global measure ($N = 24$)	Moment-to-moment measure		
	Rmean	Rpeak	PeakFreq
POST STAI State	0.55^{**}	0.42^{*}	0.56^{**}
POST STAI State - PRE STAI State	0.53^{**}	0.42^{*}	0.34

*: p <0.05, **: p <0.01

TABLE IV
PEARSON'S CORRELATION COEFFICIENT BETWEEN DIFFERENT SELF-REPORTED STRESS MEASURES AND BIO-BEHAVIORAL FEATURES. $N = 24$ FOR EDA AND SPEECH, AND $N = 19$ FOR ECG.

Modality	Feature	STAI	Rmean	Rpeak	PeakFreq
EDA	Mean SCL	0.38	0.17	0.05	0.20
	SD SCL	0.33	0.07	0.05	0.06
	SCR Amp.	0.31	0.12	-0.04	0.12
	SCR Freq.	0.31	0.36	0.36	0.33
ECG	Mean NNI	0.18	-0.12	-0.50*	-0.12
	RMSSD	0.46*	0.26	-0.02	0.07
	Mean HR	-0.20	-0.05	0.52*	0.06
	Max HR	0.45	0.45	0.69**	0.35
	LF	0.21	0.17	-0.05	-0.08
	HF	0.41	0.30	0.05	0.08
	LF-HF ratio	-0.48*	0.01	-0.05	0.12
Speech	F0	-0.05	0.04	0.13	-0.17
	Jitter	0.15	0.03	0.10	0.03
	Shimmer	0.10	-0.06	-0.04	0.01
	α ratio	0.28	0.14	0.10	0.25
	Spectral Slope	0.23	0.18	0.26	0.23

*: p <0.05, **: p <0.01

we fit a linear regression model using the POST STAI State score as the dependent variable and all the demographics and psychological trait measures as the independent variables. The weights corresponding to each of the independent variables are explored (Fig. 1) to examine how these measures contribute to each interviewee's state anxiety. From Fig. 1, it is evident that Age is the most dominant variable in estimating the POST STAI State score with older participants self-reporting higher state anxiety after the interview. An inverse relationship is observed between the POST STAI State score and the number of years since a participant left the military service. This suggests that participants who have already transitioned into civilian life reported lower state anxiety after the interview than the participants who just started transitioning. Among the psychological trait measures, both STAI Trait, and Anxiety score from SCL-90-R have positive weights, which exhibit their positive correlation with self-reported state anxiety score, which is consistent with prior work [47], [48].

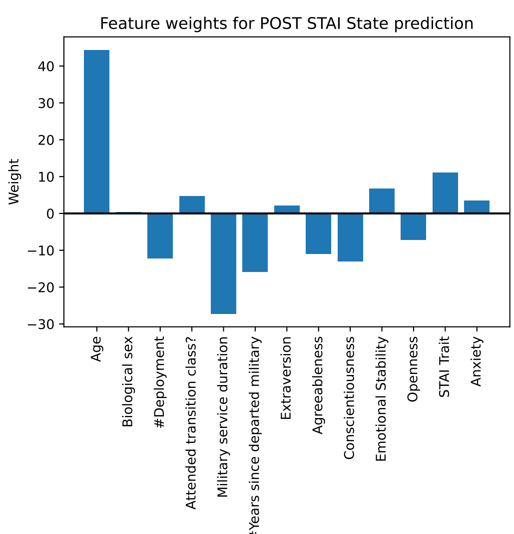

Fig. 1. Weights of demographics and individual differences measures in estimating POST STAI State score via linear regression.

TABLE V
PEARSON'S CORRELATION COEFFICIENT BETWEEN ACTUAL AND ESTIMATED STRESS MEASURES USING DIFFERENT FEATURE COMBINATION.

Feature	STAI	Rmean	Rpeak	PeakFreq
Demographic	0.01	-0.26	-0.17	-0.09
Psychological	-0.13	0.27	0.33	-0.45
Both	-0.18	0.36	0.27	-0.46
EDA	0.02	0.06	-0.18	-0.36
EDA + Demographic	0.16	0.05	0.32	0.11
EDA + Psychological	-0.02	0.40	0.48**	-0.38
EDA + Both	0.04	0.34	0.45**	-0.44
ECG	0.07	0.13	0.51*	0.02
ECG + Demographic	0.31	0.41	-0.27	-0.33
ECG + Psychological	-0.25	0.05	0.32	-0.34
ECG + Both	-0.14	0.23	0.33	-0.26
Speech	-0.09	-0.38	-0.39	0.02
Speech + Demographic	-0.15	-0.10	0.27	-0.02
Speech + Psychological	-0.09	0.02	0.56**	0.30
Speech + Both	-0.28	-0.09	0.48*	0.03

*: p <0.05, **: p <0.01

C. Estimation of Self-Reported Stress based on the Combination of Bio-behavioral and Individual Difference Measures

To answer RQ3, a leave-one-subject-out cross validation based on a linear regression model is used to estimate the stress scores from various combinations of bio-behavioral, demographic, and psychological trait measures (Section IV-E). Results of the experiment are presented in Table V. Results suggest that the estimation of the POST STAI State score depicts overall a low Pearson's r between the actual and estimated value. However, the combination of ECG features and demographic measures yield better performance ($r = 0.31$, $p = 0.19$), although the correlation is not significant. Meanwhile, the Rpeak measure is better estimated by the features than the other moment-to-moment stress measures. When combining EDA, demographic, and psychological trait measures, we are able to estimate the Rpeak with moderate correlation ($r = 0.45$, $p < 0.01$). The combination of speech and individual features also depicts strong significant correlation ($r = 0.56$, $p < 0.01$). Similar findings are observed in Table IV, where the Rpeak measure of self-reported stress depicts higher correlation with the bio-behavioral features, compared to the other moment-to-moment measures. It is interesting to note that bio-behavioral features themselves are not able to effectively estimate the stress measures, except from ECG features, which are able to estimate Rpeak with strong correlation ($r = 0.51$, $p < 0.01$). This may be a result of the low number of samples relative to the dimension of the feature space, when using the ECG features.

VI. DISCUSSION

In this work, we explored the interaction between several bio-behavioral and self-reported measures of stress during job interviews of U.S. veterans. RQ1 aims to address whether the mock interview setting used in this pilot study is able to elicit stress, which has been questioned in prior work that focuses on the analysis of job interviews [16], [17]. Results suggest that bio-behavioral measures of stress exhibit significant difference between the relaxation and interview sessions (Section V-A), indicating higher physiological reactivity during the latter.

Therefore, although the present study relies on a mock interview setting, it seems to be able to elicit stress among the participants. In contrast to the bio-behavioral measures, which depict significant differences between the relaxation and interview sessions, the global score of self-reported stress (i.e., STAI State score), collected before and after the interview, does not significantly change. A potential reason for this might be the subjectivity of the self-reported stress measures that often fail to capture the perceived stress of the study population due to the individual differences among the participants [49]. Another reason might be that the self-reports from STAI State were obtained in a PRE-POST manner, while the bio-behavioral measures are captured within the interview session.

Next, RQ2 explores the correlation between bio-behavioral measures, and the global and moment-to-moment self-reported measures of stress (Section V-B). The moderate-to-strong positive correlation between the global stress measures (i.e., STAI State score) and moment-to-moment stress measures (i.e., Rmean, Rpeak, PeakFreq) supports the internal consistency of the two. EDA features further exhibit moderate correlation with the self-reported stress indicating an association between the two. However, the trends of the ECG features seem to vary. Some ECG features depict correlation with self-reported stress along the expected direction (e.g., Max HR), while others depict a correlation in the opposite direction (e.g., RMSSD, LF-HF ratio), a finding consistent with prior work [22]. This finding might be due to the low number of samples included in the analysis of the ECG signals ($N = 17$). Another reason might be the relatively short analysis window over which the ECG measures of parasympathetic activity are computed [46]). Moreover, speech features have a low correlation with the stress measures. Since speech features tend to change quickly (i.e., at the millisecond level), aggregating them over larger windows (i.e., at the minute level) may fail to capture their temporal variation due to stressors. Finally, as part of RQ2, we explore the effect of individual differences on self-reported anxiety. The number of years since the participants were separated from the military service depicts an inverse relationship with the self-reported stress measures. This suggests that veterans who are still in transition, might feel more stressed during the interviews compared to their counterparts who have already transitioned to the civilian workforce. This result is also consistent with prior findings [5], [6]. The third research question examines the extent to which we can estimate self-reported stress by combining the different bio-behavioral measures and the individual measures (Section V-C). Results suggest that using the bio-behavioral measures alone, we can estimate self-reported state anxiety with low accuracy in most cases. This finding has been observed in prior work involving job interview analysis [17], where the authors pointed out that the inherent challenges in obtaining reliable stress labels might be a potential reason of this low accuracy. In this paper, we observe that it is relatively easier to estimate aggregates of the moment-to-moment measures of stress (*e.g.,* Rpeak) compared to the global measures of stress. A potential reason for this might be that the moment-to-moment measures capture salient stress moments within the interview, which are not easily captured by the self-report administered after the interview. Finally, augmenting the feature space of bio-behavioral measures with the interviewee's demographic and psychological characteristics tends to increase the performance of the model, which is consistent with prior work [26].

Despite the promising results, the work presented in this pilot study has some limitations. The relatively small sample size limits the generalizability of our finding and compromises the goodness of fit of the regression models (e.g., the dimensionality of the feature space is larger compared to the sample size for some of the experiments). A proper feature selection or feature reduction method might be useful in this context. So far we have worked on estimating a single score of self-reported stress that describes the entire interview session. However, as part of our future work, we will be investigating the moment-to-moment estimation and prediction of stress using the collected data. Finally, the job interview is a unique form of social interaction. Thus, analyzing the interviewee's data in tandem with the interviewer's data would allow us to investigate the interaction between the two and might improve our understanding of stressors that occur during the interview process.

VII. Conclusion

In this paper, we conducted a pilot study with 24 U.S. veterans using a mock interview setting to obtain bio-behavioral and self-reported measures of stress. The association between various measures of stress has been examined, and the effect of different individual differences has been explored. Our findings can enable the design of personalized assistive interfaces that can help military veterans succeed in civilian job interviews. As part of our future work, we plan to estimate moment-to-moment stress during the interview and use more sophisticated data-driven techniques for stress estimation.

Ethical Impact Statement

Findings from this paper can be used in designing intelligent assistive technologies that can support job candidates who seek job interview training. Despite the premise of such technologies, various ethical considerations should be taken into account. First, the small data size in combination with the heavy data bias toward male participants might affect the generalizability of the studied models. It is further important to investigate stress detection algorithms from the lens of stakeholders, who usually find it critical to understand the inner workings and decision-making processes of the algorithms. Finally, we have considered sensitive data (e.g., speech) that might reveal a user's identity. Thus, carefully-designed security protocols, that are also clearly explained to the user, should be followed for data storage and sharing.

Acknowledgment

The authors would like to thank Dr. Ani Nenkova from University of Pennsylvania for the valuable discussions in designing the study, as well as The Don & Ellie Knauss Veteran Resource and Support Center (VRSC) at Texas A&M University for their partnership in this research.

REFERENCES

[1] "Census - table results," https://data.census.gov/cedsci/table?q=veteranstid =ACSST1Y2019.S2101, (Accessed on 04/12/2022).

[2] J. E. Vespa, "Those who served: America's veterans from world war ii to the war on terror," *World War II (December 1941 to December 1946)*, vol. 485, no. 463, p. 22, 2020.

[3] C. A. Castro, S. Kintzle, and A. Hassan, "The state of the american veteran: The los angeles county veterans study," *University of Southern California center for innovation and research on veterans and military families*, 2014.

[4] "Veterans' employment challenges: Perceptions and experiences of transitioning from military to civilian life," *Prudential Financial*, 2012.

[5] J. Ahern, M. Worthen, J. Masters, S. A. Lippman, E. J. Ozer, and R. Moos, "The challenges of afghanistan and iraq veterans' transition from military to civilian life and approaches to reconnection," *PloS one*, vol. 10, no. 7, p. e0128599, 2015.

[6] S. Shepherd, D. K. Sherman, A. MacLean, and A. C. Kay, "The challenges of military veterans in their transition to the workplace: A call for integrating basic and applied psychological science," *Perspectives on Psychological Science*, vol. 16, no. 3, pp. 590–613, 2021.

[7] "Employment situation of veterans news release," https://www.bls.gov/news.release/vet.htm, (Accessed on 04/12/2022).

[8] M. C. Harrell and N. Berglass, "Employing america's veterans," *Center for a New American Security*, 2012.

[9] C. Stone and D. L. Stone, "Factors affecting hiring decisions about veterans," *Human Resource Management Review*, vol. 25, no. 1, pp. 68–79, 2015.

[10] A. H. Behzadan and T. Chaspari, "Panel Report on Veteran, Race, and Gender Bias in Tech Hiring," 2021. [Online]. Available: https://doi.org/10.18738/T8/SLYTC9

[11] M. Keeling, S. Kintzle, and C. A. Castro, "Exploring us veterans' post-service employment experiences," *Military Psychology*, vol. 30, no. 1, pp. 63–69, 2018.

[12] D. Shields, D. Kuhl, K. Lutz, J. Frender, N. Baumann, and P. Lopresti, "Mental health and well-being of military veterans during military to civilian transition: review and analysis of the recent literature," *Vancouver BC: Report prepared for Veterans Affairs Canada and the Canadian Institute for Military and Veteran Health Research*, 2016.

[13] B. Carll, "Vietnam veterans: War atrocities, social connections and assimilation," Ph.D. dissertation, 2009.

[14] C. M. Hardison, T. C. McCausland, M. G. Shanley, A. R. Saavedra, A. Clague, J. C. Crowley, J. Martin, J. P. Wong, and P. S. Steinberg, "What veterans bring to civilian workplaces: A prototype toolkit for helping private-sector employers understand the nontechnical skill developed in the military," RAND Corporation, Tech. Rep., 2017.

[15] M. Sieverding, "'be cool!': Emotional costs of hiding feelings in a job interview," *International Journal of Selection and Assessment*, vol. 17, no. 4, pp. 391–401, 2009.

[16] M. Hoque, M. Courgeon, J.-C. Martin, B. Mutlu, and R. W. Picard, "Mach: My automated conversation coach," in *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, 2013, pp. 697–706.

[17] I. Naim, M. I. Tanveer, D. Gildea, and M. E. Hoque, "Automated analysis and prediction of job interview performance," *IEEE Transactions on Affective Computing*, vol. 9, no. 2, pp. 191–204, 2016.

[18] L. Hemamou, G. Felhi, J.-C. Martin, and C. Clavel, "Slices of attention in asynchronous video job interviews," in *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 2019, pp. 1–7.

[19] L. S. Nguyen, D. Frauendorfer, M. S. Mast, and D. Gatica-Perez, "Hire me: Computational inference of hirability in employment interviews based on nonverbal behavior," *IEEE transactions on multimedia*, vol. 16, no. 4, pp. 1018–1031, 2014.

[20] L. Chen, R. Zhao, C. W. Leong, B. Lehman, G. Feng, and M. E. Hoque, "Automated video interview judgment on a large-sized corpus collected online," in *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*. IEEE, 2017, pp. 504–509.

[21] K. Anderson, E. André, T. Baur, S. Bernardini, M. Chollet, E. Chryssafidou, I. Damian, C. Ennis, A. Egges, P. Gebhard *et al.*, "The tardis framework: intelligent virtual agents for social coaching in job interviews," in *International Conference on Advances in Computer Entertainment Technology*. Springer, 2013, pp. 476–491.

[22] G. Giannakakis, D. Grigoriadis, K. Giannakaki, O. Simantiraki, A. Roniotis, and M. Tsiknakis, "Review on psychological stress detection using biosignals," *IEEE Transactions on Affective Computing*, 2019.

[23] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Transactions on intelligent transportation systems*, vol. 6, no. 2, pp. 156–166, 2005.

[24] K. Hovsepian, M. Al'Absi, E. Ertin, T. Kamarck, M. Nakajima, and S. Kumar, "cstress: towards a gold standard for continuous stress assessment in the mobile environment," in *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, 2015, pp. 493–504.

[25] H. M. Trice and J. M. Beyer, *The cultures of work organizations.* Prentice-Hall, Inc, 1993.

[26] M. Yadav, M. N. Sakib, E. H. Nirjhar, K. Feng, A. Behzadan, and T. Chaspari, "Exploring individual differences of public speaking anxiety in real-life and virtual presentations," *IEEE Transactions on Affective Computing*, 2020.

[27] S. Koldijk, M. Sappelli, S. Verberne, M. A. Neerincx, and W. Kraaij, "The swell knowledge work dataset for stress and user modeling research," in *Proceedings of the 16th international conference on multimodal interaction*, 2014, pp. 291–298.

[28] "Video conferencing, cloud phone, webinars, chat, virtual events — zoom," https://zoom.us/, (Accessed on 04/03/2022).

[29] "E4 wristband," https://www.empatica.com/en-gb/research/e4/, (Accessed on 04/04/2022).

[30] "Actiheart 5 - camntech," https://www.camntech.com/actiheart-5/, (Accessed on 04/04/2022).

[31] L. R. Goldberg, "The development of markers for the big-five factor structure." *Psychological assessment*, vol. 4, no. 1, p. 26, 1992.

[32] C. D. Spielberger, "State-trait anxiety inventory for adults," 1983.

[33] H. Calsbeek, M. Rijken, G. Henegouwen, and J. Dekker, "Factor structure of the coping inventory for stressful situations (ciss-21) in adolescents and young adults with chronic digestive disorders," *The Social Position of Adolescents and Young Adults With Chronic Digestive Disorders. Utrecht: NIVEL*, 2003.

[34] L. R. Derogatis and R. Unger, "Symptom checklist-90-revised," *The Corsini encyclopedia of psychology*, pp. 1–2, 2010.

[35] M. J. Gray, B. T. Litz, J. L. Hsu, and T. W. Lombardo, "Psychometric properties of the life events checklist," *Assessment*, vol. 11, no. 4, pp. 330–341, 2004.

[36] W. Arthur Jr, "An unproctored internet-based test of general mental ability," *A validation report. College Station, TX: Author*, 2017.

[37] J. M. Girard, "CARMA: Software for continuous affect rating and media annotation," *Journal of Open Research Software*, vol. 2, no. 1, p. e5, 2014.

[38] M. E. Dawson, A. M. Schell, and D. L. Filion, "The electrodermal system." 2017.

[39] R. J. Prineas, R. S. Crow, and Z.-M. Zhang, *The Minnesota code manual of electrocardiographic findings*. Springer Science & Business Media, 2009.

[40] "Welcome to biosppy — biosppy 0.6.1 documentation," https://biosppy.readthedocs.io/en/stable/, (Accessed on 04/06/2022).

[41] "Welcome to hrv-analysis documentation! — hrvanalysis 1.0.0 documentation," https://aura-healthcare.github.io/hrv-analysis/, (Accessed on 04/06/2022).

[42] F. Eyben, M. Wöllmer, and B. Schuller, "Opensmile: the munich versatile and fast open-source audio feature extractor," in *Proceedings of the 18th ACM international conference on Multimedia*, 2010, pp. 1459–1462.

[43] M. Benedek and C. Kaernbach, "A continuous measure of phasic electrodermal activity," *Journal of neuroscience methods*, vol. 190, no. 1, pp. 80–91, 2010.

[44] S. Laborde, E. Mosley, and J. F. Thayer, "Heart rate variability and cardiac vagal tone in psychophysiological research–recommendations for experiment planning, data analysis, and data reporting," *Frontiers in psychology*, vol. 8, p. 213, 2017.

[45] G. E. Billman, "The lf/hf ratio does not accurately measure cardiac sympatho-vagal balance," *Frontiers in physiology*, vol. 4, p. 26, 2013.

[46] T. F. of the European Society of Cardiology *et al.*, "Heart rate variability: standards of measurement, physiological interpretation and clinical use," *circulation*, vol. 93, pp. 1043–1065, 1996.

[47] B. L. Kennedy, J. J. Schwab, R. L. Morris, and G. Beldia, "Assessment of state and trait anxiety in subjects with anxiety and depressive disorders," *Psychiatric Quarterly*, vol. 72, no. 3, pp. 263–276, 2001.

[48] N. V. Ramanaiah, M. Franzen, and T. Schill, "A psychometric study of the state-trait anxiety inventory," *Journal of personality assessment*, vol. 47, no. 5, pp. 531–535, 1983.

[49] R. Rosenman, V. Tennekoon, and L. G. Hill, "Measuring bias in self-reported data," *International Journal of Behavioural and Healthcare Research*, vol. 2, no. 4, pp. 320–332, 2011.