

Stressful First Impressions in Job Interviews

Ailbhe N. Finnerty
Fondazione Bruno Kessler
CIMeC, Italy
ailbhe.finnerty@gmail.com

Skanda Muralidhar
Idiap Research Institute
EPFL, Switzerland
smuralidhar@idiap.ch

Laurent Son Nguyen
Idiap Research Institute
EPFL, Switzerland
lnguyen@idiap.ch

Fabio Pianesi
EIT Digital
FBK, Italy
fabio.pianesi@eitdigital.eu

Daniel Gatica-Perez
Idiap Research Institute
EPFL, Switzerland
gatica@idiap.ch

ABSTRACT

Stress can impact many aspects of our lives, such as the way we interact and work with others, or the first impressions that we make. In the past, stress has most commonly been assessed through self-reported questionnaires; however, advancements in wearable technology have enabled the measurement of physiological symptoms of stress in an unobtrusive manner. Using a dataset of job interviews, we investigate whether first impressions of stress (from annotations) are equivalent to physiological measurements of electrodermal activity (EDA). We examine the use of automatically extracted nonverbal cues, stemming from both the visual and audio modalities, as well as EDA stress measurements for the inference of stress impressions obtained from manual annotations. Stress impressions were found to be significantly negatively correlated with hireability ratings (i.e., individuals who were perceived to be more stressed were more likely to obtain lower hireability scores). The analysis revealed a significant relationship between audio and visual features but low predictability, while no significant effects were found for the EDA features. While some nonverbal cues were more clearly related to stress, the physiological cues were less reliable and warrant further investigation into the use of wearable sensors for stress detection.

CCS Concepts

•Human-centered computing → Ubiquitous computing; •Applied computing → Psychology; Life and medical sciences; Health informatics;

Keywords

Nonverbal Behaviour; Multimodal; Stress; Job interviews; Ubiquitous Computing

1. INTRODUCTION

In the context of organisations, stress has become the most common work-related health problem across the EU [35]. Stress is most commonly defined in the psychological literature as a state of being in which individuals perceive that they cannot adequately cope with the demands being made on them or with threats to their well-being [29]. The most common methods of assessing stress from a psychological standpoint is to use self-reported measurements of a subject's perception of their own stress levels [17, 14]. While recent research is aimed at measuring stress in everyday life without the constraints of a laboratory setting [37, 15], this task is challenging, due to the uncontrolled nature of the environment that can affect data.

Job interviews can be considered naturally stress-inducing situations. They are generally the first point of contact with future employers and play a large role in shaping our futures and, as such, are stressful life events. During a job interview it is in the applicant's best interest to manage their stress levels in order to obtain a favourable outcome (e.g., being hired for the position). The environment in which job interviews take place allows for the recording of multiple sources of data, which gives us the ability to fully investigate how stress impressions are formed during the interview process. Specifically as part of this study, we are interested in comparing the nonverbal behaviour that has been found to be relevant for successful outcomes in job interviews [32] and the nonverbal behaviour that is related to stress. This is to be able to understand whether stress, as it is experienced by the individual and perceived by others, can determine the outcome of an interview. The relationships between physiological measurements (using electrodermal activity), self-rated stress, and stress impressions (external observers ratings of a subject's stress) have to our knowledge not been examined in the past, and this present study aims at addressing one aspect of this research problem.

The overall goal of this research is to better understand the formation of first impressions in a stressful environment (i.e., job interviews), defined here as *stress impressions*, i.e., the way in which stress is perceived by others. We investigate the relationship between stress impressions and the outcome of the situation (being hired or not), the nonverbal behaviour (audio and visual) exhibited during the interview, and the physiological measurements of stress (EDA) through wearable technologies. This is illustrated in Figure 1. Towards this goal, we utilise stress impressions annotated by expert raters and a range of nonverbal behaviour, visual, audio and physiological. We examined the predictive validity (with respect to stress impressions) of (1) the nonverbal behaviour displayed by both

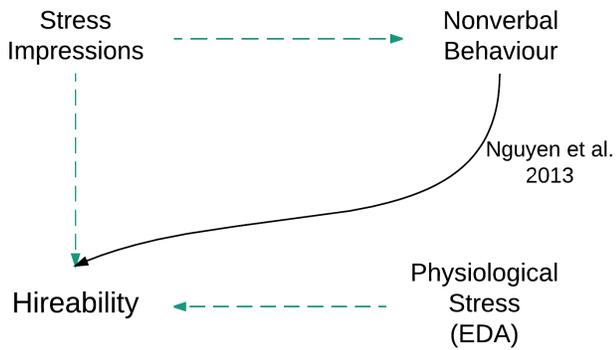


Figure 1: An illustration of the goals of this paper indicated by green dotted lines.

the job applicant and the interviewer, where the feature extraction and inference were automated and (2) the applicant’s electrodermal activity. We hypothesised the following:

- **H1** The stress impression ratings will be correlated with hireability scores.
- **H2** The stress ratings will be correlated with physiological (EDA) features.
- **H3** The visual and audio nonverbal behaviours can be used to predict the stress impressions.

Computational related works have mainly investigated stress through the use of physiological measurements; this work differs from them by investigating the formation of stress impressions (the way stress is seen by others) and its relationship to nonverbal behaviour and interview outcomes (i.e., hireability). In doing this work we aim to contribute to the growing literature on measuring stress using multimodal data sources, as we are using a combination of ubiquitous (cameras, mics) and mobile (Q sensor) sensors. Through our analysis we hope to be able to categorise specific nonverbal behaviours as stress related behaviours, as this is missing from the current literature. We also aim to contribute to the understanding of EDA data and how it can be used in unconstrained environments to reliably measure and detect stress in an individual.

2. RELATED WORK

2.1 Stress Assessment from Self-Reports

Measuring stress is a difficult task and previous research on this topic has primarily taken part in laboratory settings where the environment was strictly controlled and the stressful conditions were artificially administered [46]. The conditions were designed to stress the participant in the study; however, they did not conform to stress as it is experienced in real life, due to the fact that the conditions were static and short lived, removing the personal and natural context of the experience of stress [28]. As mentioned previously, the most common methods of assessing an individuals stress in the past were done through self-reported measurements using a variety of scales and questionnaires. However as with any form of self reporting, the measurements can be subject to bias [6] and individuals may not be aware of their own symptoms of stress, which as well as being psychological can include physiological aspects, such as blood volume pulse, increased skin temperature and galvanic skin response [50].

2.2 Stress Assessment from Wearable Devices

The increase in the availability of affective wearables [38] has changed the way in which we can measure affective states of individuals. Instead of relying on self reports we can use affective patterns, such as changes in skin conductance and heart rate, as well as changes in voice and gestures to understand the emotional states of the individual under investigation. Many recent studies have used measurements of skin conductance (SC) also known as *electrodermal activity* (EDA) to measure a range of behaviours such as stress during driving [26], sleep quality [42] and distress in individuals with disabilities.

Electrodermal Activity (EDA) is a sensitive index of the sympathetic nervous system and can be represented by the changes in skin conductance at the surface of the skin, as it reflects activity within the sympathetic axis and provides a convenient measure of assessing changes in the autonomic nervous system (ANS) [19]. Recent work has been focused on finding ways to continuously measure physiological signals from the body to determine when stress is occurring in an individual, using data obtained from both the sensors and self reports from the individual, with self reports often being used as a ground truth. This interest in wearable sensors has resulted in the development of several devices, such as, the Affectiva Q sensor [38], Empatica [23] and many others [34]. These wearable sensors have been designed to unobtrusively record physiological signals of affective states, such as arousal, and have arisen due to an interest in wearable biosensor systems for health monitoring.

2.3 Impressions in Job Interviews

Job interviews are zero acquaintance social interactions [5], as it is quite often the first time that the recruiter and the applicant will meet, leaving the recruiter with a short interaction sometimes lasting just minutes on which to base their judgement of whether to hire the applicant or not. The recruiter utilises signals from both verbal and nonverbal behaviour in order to make an informed decision on the outcome of the interaction. Even though the interactions are based on a first time meeting with a short duration, they have been found to contain enough information to accurately judge and make accurate first impressions of the individual under investigation with respect to their personality [48].

There are many factors to making a good first impression and there are many aspects of what constitutes a good first impression. Previous research has examined the effect of the short time frame of the interview, the accuracy of the judgements of recruiters, the role of self promoting behaviour [43] as well as personality and other social constructs, such as the role of nonverbal behaviour [22] on the outcome of the job interview.

With advancements in technology, it is now possible to go beyond self reported measures of internal affective states and to couple self reports with physiological data. Having both sources of data can be more telling of an individual’s internal state than their own ratings, which can be biased due to memory effects or external persuasion, such as the wording the questions being asked [45].

2.4 Related Work in Computing

A growing body of research has studied the role of nonverbal behaviour in various social interactions and settings [25, 30]. In these studies an individual’s nonverbal behaviour has been found to be predictive of social constructs, such as personality [7], interest [49] and dominance [36]. Some studies have focused on specific interactions, such as simulated job negotiations [20] and have automatically extracted audio nonverbal behaviour features, such as activity, engagement, mimicry etc, during these interactions. Other studies have focused specifically on job applicants’ affective states from

video recordings of job interviews [16] and from this have developed training systems using virtual agents that read facial expressions, as well as other aspects of nonverbal behaviour, such as speech and prosody. These studies generally use simulated situations with the aim of understanding the performance of the participant, while some studies that have used more realistic interview situations found that nonverbal immediacy was linked to the recruiter hiring decision [22].

Auditory cues of stress that can be perceived by the listener vary from obvious to subtle, these cues are both consciously and unconsciously monitored by the speaker. There are auditory changes that occur to speech due to stress, these changes are observable to the speaker, such as a variations in pitch and energy, and they are equally observable to the listener, leading to the phenomena of social masking [13]. The verbal indicators of stress can be identified easily, like stuttering, repetition, and tongue-slip, but the nonverbal cues are more difficult to conceal, as they are caused by the autonomous nervous system (ANS). Listeners can perceive the auditory cues in the speaker’s speech even if these are not obvious. These cues are perceived not merely based on the acoustic signal but also paralinguistics obtained in the context of the conversation and their experience [44].

3. INTERVIEW DATASET

In this study, we used a subset of the job interviews dataset first used in [32]. This dataset is composed of recorded job interviews of 54 individuals. They were interviewing for a marketing position to recruit participants for psychological studies at a university. The majority of the applicants were students, with an average age of 24 years (st dev = 5.68 years), with more female (45) than male (17) applicants.

Structured interviews, consisting of eight questions, were used to collect the data to ensure that the procedure was consistent across all participants. All interviews were conducted by a researcher in organisational psychology and a snapshot of the experimental setup can be seen in Figure 2. During the interview the applicants were asked for 1) a short self-presentation, 2) their motivation for applying for the job, and 3) their opinion on the importance of scientific research. They were also asked for their past experiences where 4) communication skills and 5) persuasion skills were required. Applicants were also asked about their past experiences of 6) conscientious or serious work, 7) where a stressful situation was correctly managed, and finally 8) some strong and weak points about themselves. Applicants then filled in a range of psychometric questionnaires to assess 1) personality, using the Big Five personality model [18], 2) intelligence [4] and, 3) communication and persuasion [41].

The interviews were recorded using two front facing 1280 x 960 cameras, facing the interviewer and the applicant, recording synchronously at 26.6 frames per second. Audio was recorded at 48kHz using a Microcone device [3], which consists of a microphone array that automatically detects speaking turns in addition to recording the audio. The Microcone was placed in the centre of the table. A detailed description of the data collection can be found in [32].

In addition to the audio and video data, the participants wore a

Q sensor, which was used to record electrodermal activity (EDA) [39] data. EDA can be used as a measurement of arousal which, depending on the context of a situation, can be a measure of negative affect or whether an individual is excited or stressed.

3.1 Annotations

First impressions of stress was rated by three independent raters who watched the videos of all participants. Specifically, they were asked to rate “How stressed do you think the applicant is?” on a



Figure 2: A snapshot of the experimental recording setup in the lab (top). Images of the interviewer (bottom left) and the job applicant (bottom right) as recorded by the HD cameras. Image taken from [32].

five-point Likert scale (1=‘not at all’ to 5=‘very much’) for each of the eight questions. The raters were also asked to provide an overall measure of their impression of how stressed the applicant was for the entire duration of the interview. All three raters were masters students in Psychology.

The agreement between the raters was calculated as the Intraclass Correlation Coefficient [47], a standard measure of agreement used in psychology and social computing, for each of the eight interview questions and the overall measure of stress. The $ICC(2, k)$ values were above 0.50 for each of the questions and for the overall rating of stress, which we use as part of our experiments, the agreement was 0.67 ($p < .001$).

The hireability impressions was rated by professional recruiters. They watched the interviews and rated the participants on a scale from 0 to 100 percent (with increments of 10%) of how likely they were to hire a particular applicant and how suited they were to the position ($M = 59.45$, $SD = 19.59$). Inter-rater agreement between the raters was $ICC(1, 1) \in [.32, .52]$ and $ICC(1, k) \in [.59, .77]$, using the intraclass correlation coefficient.

4. FEATURES

In this work, we automatically extracted features from multiple modalities, i.e., EDA, audio, and visual, and fused them together into a multimodal feature set. Previous literature has shown that an interviewer’s nonverbal behaviour can also have an impact on the outcome of the interview [32, 21]. Thus, we extracted audio and visual cues automatically from both the applicant and the interviewer.

4.1 Audio Features

Speech was segmented into applicant and interviewer automatically through the use of the Microcone, a commercial array of microphones that performs speaker segmentation via sound source localisation. Then, the following prosodic (MPEG-7 & MFCC) and speaking activity features was extracted.

- **MPEG-7:** This provides a framework for describing audio signals using signal-based criteria that can be automatically extracted at any desired temporal (or spectral) resolution [40]. This framework of audio descriptors provides a comprehensive set of features. In this work, we utilised the MATLAB[®] code for MPEG-7 [1]. The features were calculated using a 100ms window with 50% overlap instead of the classic 40ms window, so as to optimise the amount of samples for the cal-

ulation of different features. Various statistics, such as mean, median, standard deviation, min and max were computed and used as features.

- **Mel-Frequency Cepstral Coefficients:** MFCC is a common acoustic feature used for speech emotion recognition. It stems from the fact that MFCCs are the most popular spectral based parameters that are used for spectral representation of speech signal. These have traditionally been used in various speech applications, including stress speech recognition in [10, 11]. In this work, we determine the MFCC coefficient values using the following standard parameters: 24 cepstral coefficients, 13 linear filter and 27 log filters with a window size of 40ms and overlap of 10ms.
- **Speaking time:** Total speaking time of the applicant obtained by summing all the individual speaking segments. This was then normalised by the interview duration.
- **Speaking turn:** Number of speech segments longer than two seconds.
- **Pauses:** Number of gaps in speech less than two seconds in duration. This number was then normalised by the interview duration.
- **Short utterances:** These are speaking segments of duration less than two seconds. These were again normalised with the interview duration.
- **Audio back-channelling:** These are audio event where a person produced a short utterance while the other was speaking.

4.2 Visual Features

Visual features were automatically extracted for both the applicant and interviewer using the methods proposed in [32].

- **Head nods:** Vertical up-and down movements of the head, rhythmically raised and lowered. Heads nods were automatically extracted using the methods proposed in [31]. From the detected head nods, the number of nods and total nodding time was recorded and normalized by average interview duration.
- **Overall visual motion:** Visual movement displayed by the applicant and interviewer during the job interview. This acts as an indicator of kinetic expressiveness [9]. From this feature, we extracted the following statistical features: mean, median, standard deviation, minimum, maximum, entropy, quartiles, and centre of gravity.
- **Head visual motion:** This is the amount of head motion displayed by a person. This feature is based on the parametric optical flow estimation method described in [33]. A number of statistics were computed (e.g., mean, standard deviation, zero-crossing rate, entropy) for the head region.
- **Visual back-channelling:** Nods by a person while the other speaks.

4.3 EDA Features

The Affectiva Q sensor TM [2] was used as a measurement of the EDA of the interview applicants. We used the Q sensor as it is an unobtrusive wearable sensor in a bracelet form that is worn on the wrist and allows the wearer free movement without any disturbances. The device allows different sampling frequencies of 8, 16 and 32 Hertz (Hz) and provides a measure of EDA in microsiemens (μS).

The data was preprocessed, and in addition to computing basic EDA data statistics, we computed high level features which are more informative of the affective states of the individual. We used the Ledalab software designed specifically for the pre-processing and feature extraction of EDA data. The data was processed using the Continuous Decomposition Analysis (CDA) option [8] which is

designed to separate out the EDA raw data into continuous phasic and tonic components and computes several standard measures of phasic EDA.

These features were calculated for each event and are described as follows (also see Table. 1).

- **AmpSum:** sum of the skin conductance response (SCR) amplitudes of the significant SCRs (above the threshold), measured in microsiemens μS .
- **Latency:** time period from the beginning of the event to the peak of the first significant SCR, measured in seconds (s).
- **nSCR:** number of significant skin conductance responses above a predetermined threshold ($0.01\mu S$), measured as a whole number.
- **SCR:** average phasic activity, measured in microsiemens μS .
- **ISCR:** cumulative phasic activity, which is the same as the value of average phasic activity (SCR) multiplied by the size of the response window, the duration of the interview question (seconds), measured in microsiemens μS .

Table 1: Basic Statistics for EDA features

	AmpSum	Latency	nSCR	SCR	ISCR
Mean	0.266	7.35	5.389	0.038	2.085
Median	0.213	5.13	5.13	0.036	1.618
St Dev	0.225	6.01	2.924	0.036	1.940
Max	1.081	29	14.625	0.215	10.638
Min	0	0	0	0.002	0.092

5. EXPERIMENTS

In this section we describe the analysis used to compare and contrast the different data measurements used to understand stress.

5.1 Correlation Analysis

A Spearman rank correlation analysis was done on the data to compare the stress impressions and the outcome of the job interview hireability scores (**H1**). Correlations were then done to compare the stress impressions, the hireability and the various features extracted from the dataset. These included the nonverbal (visual and audio) and physiological features (EDA) to investigate our second hypothesis (**H2**). Spearman correlation was used under the assumption that the data was not normally distributed. For the analysis we used the average of the stress impressions provided by the three raters, the basic averaged values of nonverbal (audio and visual) features, and the EDA stress features calculated using the software Ledalab.

5.2 Regression Analysis

Following the initial exploratory analysis, we proposed and evaluated a computational framework for the automatic inference of stress impressions from nonverbal and physiological behaviours with the stress impressions to investigate our third hypothesis (**H3**). We defined the inference task as a regression problem, i.e. predicting the Stress ratings from nonverbal behavioural cues. To this end, the data was pre-processed by a speaker independent Z-score normalisation to transform data into unity variance and zero mean. As a second step, the following dimensionality reduction techniques were evaluated.

- **Principle Component Analysis (PCA):** This is a classic procedure that orthogonally transforms the original n coordinates of a data set into a new set of n coordinates. The result of the transformation being that the first principal component has the largest possible variance. In this study, the number of principal components was set such that 99.9% of the variance could be explained by the model.

- Low pVal (pVal): this method assumes that the relevant information is contained in the features that are significantly correlated with the social variables. We selected features from the training set with $p < .05$.
- All features (all). In order to test the improvement of the dimensionality reduction step, we also tested the case of taking all features as predictors for the regression step.

We then evaluated two regression techniques, Ridge regression (Ridge) and random forest (RF) to predict stress impressions. For these tasks, leave-one-interview-out cross validation and 10-fold inner cross validation were used. We used two metrics to evaluate the performance of the automatic prediction models: the root-mean-square error (RMSE) and the coefficient of determination (R^2). Both metrics are widely used measures in psychology and social computing. As the baseline regression model, we took the average stress impression scores as the predicted value. The RMSE is the difference between a model's predicted values and the values actually observed. The coefficient of determination (R^2) is based on the ratio between the mean squared errors of the predicted values, obtained using a regression model and the baseline-average model.

6. RESULTS

In this section, results of correlation and regression analysis are presented. The results will be interpreted and discussed further in the subsequent section.

6.1 Correlation Analysis

Stress Impressions, Hireability and Traits. The hireability scores, as measured by expert recruiters, were found to be significantly negatively correlated with the stress impressions, such that the individuals who appeared more stressed during the interview were less likely to have high hireability scores.

Further analysis was done on personality traits, from self reported measurements, personality traits as perceived by the recruiter, and stress impressions. The results (Table. 2) show that two personality traits were negatively correlated with stress impressions, Openness to Experience and Conscientiousness. For the traits as perceived by the recruiter, Conscientiousness was negatively correlated with stress impressions while Neuroticism was positively correlated. Other findings from the analysis reveal that those who were rated as being more stressed were less likely to be perceived as being persuasive, being able to successfully manage stressful situations or giving a good overall impression.

As well as correlating the stress impressions with the trait behaviour from the questionnaires, we report the correlations with the hireability scores in the tables. As being rated more stressed was correlated with hireability we were interested in whether the features that were correlated with stress were also correlated with hireability. Results indicate that features that are negatively correlated to stress impressions were positively correlated to hireability scores.

Stress Impressions and Visual Features. Stress impressions were significantly correlated with very few of the visual nonverbal features. The features that were significantly correlated with the stress impressions were negative, while the same features were positively correlated with the hireability scores. Interestingly the significant features were only related to the interviewer and not to the applicant. They were the number of times and the amount of time that the interviewer nodded their head, the number of visual back-channelling events, and the time spent back channelling. The full set of results can be seen in Table. 3.

Table 2: Trait Behaviour Correlations with Stress Impressions and Hireability Scores. Note that ** is significant at value $p < .001$, * is significant at $p < .05$

	Stress	Hiring
Personality Traits		
Extraversion	-0.13	0.44*
Openness to Experience	-0.43**	0.16
Neuroticism	0.21	-0.15
Agreeableness	-0.03	0.09
Conscientiousness	-0.28**	0.25
Impressions of Personality		
Extraversion	-0.24	0.43*
Openness to Experience	-0.02	0.26
Neuroticism	0.38**	-0.12
Agreeableness	-0.13	0.41**
Conscientiousness	-0.43**	0.34**
Recruiter Scores		
CRED Communication	-0.04	0.43*
CRED Persuasion	-0.16	0.27**
CRED Conscientiousness	-0.29**	0.31**
CRED Stress Management	-0.14	0.21
Impressions of Coders		
Persuasion	-0.60*	0.80*
Conscientiousness	-0.48*	0.63*
Stress Management	-0.53*	0.71*
General Impression	-0.50*	0.89*

Table 3: Visual Nonverbal Feature Correlations with Stress Ratings and Hireability Scores. Note that ** is significant at value $p < .001$, * is significant at $p < .05$

	Stress	Hiring
Applicant Number of Nods	0.03	0.18
Applicant Nodding Time	-0.03	0.22
Interviewer Number of Nods	-0.40*	0.45**
Interviewer Nodding Time	-0.52**	0.54**
Applicant Number of Visual Backchannel Events	-0.01	0.22
Applicant Visual Backchanneling Time	-0.04	0.26
Interviewer Number of Visual Backchannel Events	-0.48**	0.62**
Interviewer Visual Backchanneling Time	-0.55**	0.65**

Stress Impressions and Audio Features. Unlike the visual features, there were many more significant correlations between the audio features and stress impressions. For the audio feature analysis, the majority of the significant features were from speech produced by the applicant and very few related to the interviewer. It was found that short utterances, the amount of silent events, and time spent silent were positively correlated with stress impressions, as well as the applicant number of turns speaking and when the interviewer was speaking and nodding at the same time. Stress impressions were negatively correlated with interviewer mean speaking energy, for both applicant and interviewer mean voiced rate as well as applicant speaking time. These were some of the main findings from the analysis, the complete list of all significant correlations can be seen in Table. 4.

Regarding MFCCs, the spectral audio features that were extracted were found to have both positive and negative correlations with the stress impressions. Mean MFCC and Cov Diff MFCC were both negatively correlated with stress impressions, while Mean ASS and Mean ASP were positively correlated with stress impressions. The results can be seen in Table. 5.

Stress Impressions and EDA. The stress impressions were not correlated with any of the EDA features. The correlations were

Table 4: Audio Feature Correlations with Stress Ratings and Hireability Scores. Note that ** is significant at value $p < .001$, * is significant at $p < .05$

	Stress	Hiring
Applicant Number of Short Utterances	0.35*	-0.46**
Applicant Short Utterances Time	0.38*	-0.46**
Interviewer Number of Short Utterances	0.00	-0.29*
Interviewer Short Utterances Time	0.02	-0.29*
Applicant Number of Turns	0.37*	-0.51**
Applicant Speaking Time	-0.36*	0.48**
Applicant Average Turn Duration	-0.37*	0.54**
Applicant Maximum Turn Duration	-0.54**	0.55**
Interviewer Nodspeak Time	0.27*	-0.18
Number of Silent Events	0.33*	-0.55**
Time Silent	0.38*	-0.58**
Applicant Mean Energy	-0.25	0.08
Applicant Energy Standard Deviation	-0.17	0.00
Applicant Median Energy	-0.30*	0.40*
Applicant Energy Lower Quartile	-0.38*	0.51**
Applicant Energy Upper Quartile	-0.22	0.09
Interviewer Mean Energy	-0.31*	0.05
Interviewer Energy Standard Deviation	-0.29*	0.00
Interviewer Median Energy	-0.34*	0.24
Interviewer Energy Lower Quartile	-0.36*	0.30*
Interviewer Energy Upper Quartile	-0.29*	0.03
Applicant Mean Voiced Rate	-0.36*	0.54**
Applicant Voiced Rate Standard Deviation	-0.29*	0.47**
Applicant Median Voiced Rate	-0.32*	0.39*
Interviewer Mean Voiced Rate	-0.27*	0.42*

Table 5: Spectral Audio Feature Correlations with Stress Ratings and Hireability Scores. Note that ** is significant at value $p < 0.001$, * is significant at $p < 0.05$

Audio Spectral Features	Stress	Hiring
Mean Fundamental Frequency	0.05	-0.47**
Cov Fundamental Frequency	-0.03	0.43*
Mean Audio Spectral Centroid	-0.23	0.50**
Mean Audio Spectrum Spread	0.27*	-0.31*
Cov Audio Spectral Basis	-0.02	0.28*
Mean Audio Spectral Projection	0.36*	-0.49**
Mean MFCC	-0.36*	0.50**
Cov Diff MFCC	-0.35*	0.19

run for the overall stress impression scores and the EDA features averaged across the entire length of the interview. The analysis was run multiple times to control for any errors.

The findings overall did not support our initial hypothesis, however there are a variety of reasons that could have contributed to these results and will be assessed further in the next section.

6.2 Regression Results

The results of the experiments are shown in Table. 6. The table shows the performance of the different models for the inference of stress impressions. Results obtained for the stress impression variable were significantly better than the baseline-average model for ridge regression ($p < .05$), using the pValue dimensionality reduction technique. Performance of random forest and ridge regression using all nonverbal features were significantly better than the baseline.

The best prediction results for stress impressions was obtained using all features from the applicant’s nonverbal behaviour and random forest ($R^2 = 0.195$). While the second best results was obtained using visual features with pVal dimensionality reduction technique and ridge regression ($R^2 = 0.190$). Interviewer nonverbal

features were found to be predictive of stress impressions using all features with ridge regression ($R^2 = 0.187$).

7. DISCUSSION

From the correlation analysis, we found that stress impressions were negatively correlated with the hireability scores for the applicants, supporting our first hypothesis **H1**. The correlation suggests that the more stressed that the applicant appears to an observer, the less hireable the appear. As well as the hireability score, we were interested in whether the stress impressions were related to other traits that are considered important in a successful applicant. We found that those who were rated as more stressed were less likely to be conscientious, for both self assessments and the score given by an independent rater [12]. Those with higher stress scores were also rated as being less able to manage stressful situations and as being less persuasive, which were both desired traits for the position. They also gave a less positive general impression than those who were rated low in stress impressions, which is an important factor in job interview outcomes [24].

The correlations between the stress impressions and the EDA features revealed no significant results and therefore did not support our second hypothesis **H2**. There are a number of reasons why we believe that this occurred. 1) EDA data is a continuous signal and the time taken to change from stressed to not stressed can take longer than the change from one interview question to the next, 2) masking might be playing a role, in which some participants are good at hiding their stress, 3) EDA is technically a measurement of arousal, which strictly speaking is not stress, and finally 4) the ratings are a measurement of the outward expression of stress while the EDA data is a measurement of the internal physiological signs of stress which may not have any impact on the outward expression of the interviewees behaviour. These possible explanations could be investigated as part of future work.

Correlations between a few visual features revealed that the non-verbal behaviour of the interviewer was negatively significantly correlated with the stress impressions. This could be because visual feedback from the interviewer can serve to reassure interviewees during the process, which can increase confidence during the interview process [27]. Applicant audio cues were more significantly correlated with the stress impressions than the interviewers. Features such as number of short utterances and when there was silence during the interview, could suggest that the applicant was unsure of how to answer and less fluid in speaking, which could have appeared as stressed during these silences. From the negative correlations, we found that speech features such as speaking time, speaking energy, as well as more time spent speaking was correlated with being less stressed. This could be that as the speech was more fluent the applicant appeared more confident and performed better. This can also be seen in the corresponding positive correlations between the hireability scores for the same features.

The regression results show evidence that visual features were better predictors of stress impressions than audio features. While there was some evidence of nonverbal features being able to predict stress impressions we can determine that our third hypothesis **H3** was only partially supported by the results. While there were many correlations between the nonverbal features and the stress ratings, their predictive value was lower than expected and they only partially supported our hypothesis. Regression results for stress impressions were lower for other constructs, but up to 20% of the variance can be explained by nonverbal behaviour. The use of other types of cues (e.g., gaze, facial expressions) could be investigated in future work.

The consistent relationship between the stress impressions and the hireability scores was an interesting finding in this study. Not only

Table 6: Performance of various regression methods for various features to infer stress impression, $N = 54$

	NVB Applicant		NVB Interviewer		Audio		Visual		EDA	
	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
Baseline-Avg	0.000	0.167	0.000	0.167	0.000	0.167	0.000	0.167	0.000	0.167
All-Ridge	0.156	0.152	0.187	0.149	-0.017	0.168	0.135	0.155	-0.128	0.177
All-RF	0.195	0.148	0.085	0.158	0.042	0.163	0.103	0.158	-0.147	0.179
Pval-Ridge	0.110	0.156	0.179	0.150	-0.019	0.168	0.190	0.150	-0.113	0.176
Pval-RF	0.077	0.159	0.091	0.158	0.105	0.158	0.060	0.162	-0.417	0.199
PCA-Ridge	0.159	0.152	0.186	0.149	-0.017	0.168	0.049	0.163	-0.115	0.176
PCA-RF	0.053	0.161	0.060	0.160	0.011	0.166	0.004	0.167	-0.022	0.169

did we find that higher stress impression scores were related to lower scores on hireability, but we saw some evidence in the correlations between the hireability scores, the stress impressions and a few of the of nonverbal features under investigation. Throughout the analysis we found that when the stress impressions were negatively correlated, the hireability scores were positively correlated. This is an indication that the behavioural cues exhibited by the applicants were linked to the appearance of being stressed and thus interpreted by the rater. The link between the two variables leads us to believe that we have found some indications of which behaviours lead to stress impressions being formed, which can impact the outcome of the hiring process during job interviews. However we did not find the expected link between the stress impressions and the EDA features from the wearable device which we expected to find.

8. CONCLUSION

We conducted a study to investigate stress in an environment where a first impression is based on a very short interaction. More specifically we investigated the *stress impressions* that the applicants portrayed to external raters based on their performance during a job interview. We found that applicants with higher stress ratings scored lower on hireability, showing evidence that stress as perceived by external observers had a negative impact on performance during a job interview and it's outcome.

As stress is linked to multiple modalities of nonverbal behavioural cues, extracted from visual, audio and physiological data, we examined whether there was a link between these features and the stress impressions. We found that visual cues from the interviewer were related to lower stress impression ratings, leading us to believe the positive visual feedback is linked to positive reinforcement causing the applicant to exhibit fewer signs of stress. Audio features from the applicant showed evidence that more fluency, energy and time spent speaking during the interview was linked to lower stress impressions, whereas shorter utterances and more time spent in silence was linked to higher stress impressions. The physiological features from the EDA data were not found to be correlated with stress impressions.

Through regression experiments we found that, while there were some positive results from the analysis, there was low predictability of the stress ratings using just the nonverbal behaviour features (up to 20% of variance). Using visual features from the interviewer and audio features from the applicant produced the best results.

While the results of the experiments had low predictability, there is some evidence that there are nonverbal features which can be used in stress impression prediction in future studies. We believe that the experiments here should be extended to further understand which nonverbal behaviour is present during stressful situations. Knowing what are the signs of stress can help in understanding and measuring stress in individuals and facilitate the reduction of stress as a health concern. Further, applications could use this knowledge

of stress signals, as rated by others, in automated feedback systems to help applicants improve their interviewing techniques by better managing their stress.

This work studies human interactions though multiple modalities. We found that stress is more than physiological signals and the way stress is interpreted by others might differ from the stress one can experience. Researchers should therefore be careful when using signals like EDA to measure stress. Future work could involve the development of a system to provide meaningful behavioural feedback based on stressed nonverbal behaviour.

9. ACKNOWLEDGMENTS

This work was supported by the SNSF through the Ubimpressed project. We thank the student annotators for their work, and Denise Frauendorfer and Marianne Schmid Mast (University of Lausanne) for help and discussions.

10. REFERENCES

- [1] <http://mpeg7.doc.gold.ac.uk/>.
- [2] <http://qsensor-support.affectiva.com/>.
- [3] <http://www.dev-audio.com/products/microcone/>.
- [4] <http://www.wonderlic.com/assessments/ability/cognitive-ability-tests>.
- [5] N. Ambady, M. Hallahan, and R. Rosenthal. On judging and being judged accurately in zero-acquaintance situations. *J. of Personality and Social Psychology*, 69(3), 1995.
- [6] H. J. Arnold and D. C. Feldman. Social desirability response bias in self-report choice situations. *Academy of Management J.*, 24(2), 1981.
- [7] L. M. Batrinca, N. Mana, B. Lepri, F. Pianesi, and N. Sebe. Please, tell me about yourself: automatic personality assessment using short self-presentations. In *Proc. ACM ICMI*, 2011.
- [8] M. Benedek and C. Kaernbach. A continuous measure of phasic electrodermal activity. *J. of neuroscience methods*, 190(1), 2010.
- [9] J.-I. Biel and D. Gatica-Perez. The youtube lens: Crowdsourced personality impressions and audiovisual analysis of vlogs. *IEEE Trans. on Multimedia*, 15(1), 2013.
- [10] H. Bořil, P. Boyraz, and J. H. Hansen. Towards multimodal driver's stress detection. In *Digital signal processing for in-vehicle systems and safety*. Springer, 2012.
- [11] S. E. Bou-Ghazale and J. H. Hansen. A comparative study of traditional and newly proposed features for recognition of speech under stress. *Speech and Audio Processing, IEEE Trans. on*, 8(4), 2000.
- [12] A.-M. Brouwer, M. van Schaik, J. van Erp, and H. Korteling. Neuroticism, extraversion and stress: Physiological correlates.

In *Affective Computing and Intelligent Interaction (ACII)*, 2013 Humaine Association Conf. on. IEEE, 2013.

- [13] D. A. Cairns and J. H. Hansen. Nonlinear analysis and classification of speech under stressed conditions. *The J. of the Acoustical Society of America*, 96(6), 1994.
- [14] C. S. Carver, M. F. Scheier, and J. K. Weintraub. Assessing coping strategies: a theoretically based approach. *J. of personality and social psychology*, 56(2), 1989.
- [15] E. Ceja, V. Osmani, and O. Mayora. Automatic stress detection in working environments from smartphone's accelerometer data: A first step. 2015.
- [16] L. Chen, S.-Y. Yoon, C. W. Leong, M. Martin, and M. Ma. An initial analysis of structured video interviews by using multimodal emotion detection. ACM, 2014.
- [17] S. Cohen, T. Kamarck, and R. Mermelstein. A global measure of perceived stress. *J. of health and social behavior*, 1983.
- [18] P. T. Costa and R. R. McCrae. Normal personality assessment in clinical practice: The neo personality inventory. *Psychological assessment*, 4(1), 1992.
- [19] H. D. Critchley. Book review: electrodermal responses: what happens in the brain. *The Neuroscientist*, 8(2), 2002.
- [20] J. R. Curhan and A. Pentland. Thin slices of negotiation: predicting outcomes from conversational dynamics within the first 5 minutes. *J. of Applied Psychology*, 92(3), 2007.
- [21] T. DeGroot and J. Gooty. Can nonverbal cues be used to make meaningful personality attributions in employment interviews? *J. of Business and Psychology*, 24(2), 2009.
- [22] D. Frauendorfer and M. S. Mast. The impact of nonverbal behavior in the job interview. *The Social Psychology of Nonverbal Communication*, 2014.
- [23] M. Garbarino, M. Lai, D. Bender, R. W. Picard, and S. Tognetti. Empatica e3a wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition. In *Wireless Mobile Communication and Healthcare (Mobihealth)*, 4th Int Conf. on. IEEE, 2014.
- [24] W. L. Gardner and M. J. Martinko. Impression management in organizations. *J. of management*, 14(2), 1988.
- [25] D. Gatica-Perez. Automatic nonverbal analysis of social interaction in small groups: A review. *Image and Vision Computing*, 27(12), 2009.
- [26] J. A. Healey and R. W. Picard. Detecting stress during real-world driving tasks using physiological sensors. *Intelligent Transportation Systems, IEEE Trans. on*, 6(2), 2005.
- [27] M. E. Hoque, M. Courgeon, J.-C. Martin, B. Mutlu, and R. W. Picard. Mach: My automated conversation coach. In *Proc. ACM Int joint Conf. on Pervasive and Ubiquitous Computing*. ACM, 2013.
- [28] R. Kocielnik, N. Sidorova, F. M. Maggi, M. Ouwerkerk, and J. H. Westerink. Smart technologies for long-term stress monitoring at work. In *Computer-Based Medical Systems (CBMS)*, 2013 IEEE 26th Int Symposium on. IEEE, 2013.
- [29] R. S. Lazarus. Psychological stress and the coping process. 1966.
- [30] H. Lu, D. Frauendorfer, M. Rabbi, M. S. Mast, G. T. Chittaranjan, A. T. Campbell, D. Gatica-Perez, and T. Choudhury. Stresssense: Detecting stress in unconstrained acoustic environments using smartphones. In *Proc. ACM Conference on Ubiquitous Computing*, 2012.
- [31] L. Nguyen, J.-M. Odobez, and D. Gatica-Perez. Using self-context for multimodal detection of head nods in face-to-face interactions. In *Proc. ACM Int Conf. on Multimodal interaction*, 2012.
- [32] 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 Trans. on Multimedia*, 16(4), 2014.
- [33] J.-M. Odobez and P. Bouthemy. Robust multiresolution estimation of parametric motion models. *J. of visual communication and image representation*, 6(4), 1995.
- [34] A. Pantelopoulos and N. G. Bourbakis. A survey on wearable sensor-based systems for health monitoring and prognosis. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Trans. on*, 40(1), 2010.
- [35] P. Paoli. *Third European survey on working conditions 2000*. Office for official publications of the European Communities, 2001.
- [36] F. Pianesi, N. Mana, A. Cappelletti, B. Lepri, and M. Zancanaro. Multimodal recognition of personality traits in social interactions. In *Proc. ACM ICMI*, 2008.
- [37] R. W. Picard. Measuring affect in the wild. In *Affective Computing and Intelligent Interaction*. Springer, 2011.
- [38] R. W. Picard and J. Healey. Affective wearables. *Personal Technologies*, 1(4), 1997.
- [39] M.-Z. Poh, N. C. Swenson, and R. W. Picard. A wearable sensor for unobtrusive, long-term assessment of electrodermal activity. *Biomedical Engineering, IEEE Trans. on*, 57(5), 2010.
- [40] S. Quackenbush and A. Lindsay. Overview of mpeg-7 audio. *IEEE Trans. on Circuits and Systems for Video Technology*, 11(6), 2001.
- [41] R. E. Riggio. Assessment of basic social skills. *J. of Personality and social Psychology*, 51(3), 1986.
- [42] A. Sano and R. W. Picard. Toward a taxonomy of autonomic sleep patterns with electrodermal activity. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual Int Conf. of the IEEE*, 2011.
- [43] M. Schmid Mast, D. Frauendorfer, and L. Popovic. Self-promoting and modest job applicants in different cultures. *J. of Personnel Psychology*, 10(2), 2011.
- [44] M. Schreuder. *Prosodic processes in language and music*. PhD thesis, 2006.
- [45] N. Schwarz. Self-reports: how the questions shape the answers. *American psychologist*, 54(2), 1999.
- [46] S. C. Segerstrom and G. E. Miller. Psychological stress and the human immune system: a meta-analytic study of 30 years of inquiry. *Psychological bulletin*, 130(4), 2004.
- [47] P. E. Shrout and J. L. Fleiss. Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin*, 86(2), 1979.
- [48] A. Vinciarelli and G. Mohammadi. A survey of personality computing. *Affective Computing, IEEE Trans. on*, 5(3), 2014.
- [49] B. Wrede and E. Shriberg. Spotting "hot spots" in meetings: human judgments and prosodic cues. In *INTERSPEECH*, 2003.
- [50] J. Zhai and A. Barreto. Stress detection in computer users based on digital signal processing of noninvasive physiological variables. In *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual Int Conf. of the IEEE*, 2006.