Examining the Influence of Personality and Multimodal Behavior on Hireability Impressions

Harshit Malik* and Hersh Dhillon*
IIT Ropar, Punjab, India
harshit.malik999@gmail.com,hershdhillon23@gmail.com

Ravikiran Parameshwara
University of Canberra, Australia
ravikiran.parameshwara@canberra.edu.au

Roland Goecke
University of Canberra, Australia
roland.goecke@ieee.org

Ramanathan Subramanian
University of Canberra, Australia
ram.subramanian@canberra.edu.au

ABSTRACT

While personality traits have been traditionally modeled as behavioral constructs, we novelly posit *job hireability* as a *personality construct*. To this end, we examine correlates among personality and hireability measures on the *First Impressions Candidate Screening* dataset. Modeling hireability as both a discrete and continuous variable, and the *big-five* OCEAN personality traits as predictors, we utilize (a) multimodal behavioral cues, and (b) personality trait estimates obtained via these cues for hireability prediction (HP). For each of the *text*, *audio* and *visual* modalities, HP via (b) is found to be more effective than (a). Also, superior results are achieved when hireability is modeled as a continuous rather than a categorical variable. Interestingly, eye and bodily visual cues perform comparably to facial cues for predicting personality and hireability. Explanatory analyses reveal that multimodal behaviors impact personality and hireability impressions: *e.g.*, Conscientiousness impressions are impacted by the use of *positive adjectives* (verbal behavior) and *eye movements* (non-verbal behavior), confirming prior observations.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**;
• **Computing methodologies → Neural networks**; **Natural language processing**; **Computer vision tasks**.

KEYWORDS

Hireability, Personality traits, Multimodal, Behavioural cues, Regression, Classification

ACM Reference Format:
Harshit Malik* and Hersh Dhillon*, Ravikiran Parameshwara, Roland Goecke, and Ramanathan Subramanian. 2023. Examining the Influence of Personality and Multimodal Behavior on Hireability Impressions. In *Indian Conference on Computer Vision, Graphics and Image Processing (ICVGIP '23), December 15–17, 2023, Rupnagar, India.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3627631.3627658

1 INTRODUCTION

Behavioral cues such as *eye movement*, *gestural*, *facial, gazing* and *neural patterns* have been employed for cognition and emotion prediction [18, 21], personality trait estimation [2, 24], depression detection [5], gender prediction [4] and cognitive load estimation [3, 16]. Recently, multimodal behavioral cues have been utilized for predicting job interview outcomes [6, 9, 17]. Given the many applications received by top companies on a daily basis [19], there has been an increased push for employing *artificial hiring agents* (AHAs) to recruit candidates; the rationale is that AHAs assess a large pool of candidates in the early rounds, while professional recruiters interview the most promising ones in the later stages.

To make the recruitment process transparent and trustworthy, AHAs need to *justify* their decisions with *explanations*. A handful of works have employed both verbal and non-verbal behavioral cues to predict a candidate's apparent ***hireability***, *i.e.*, the suitability of a candidate to be invited for interview later; hireability prediction (HP) is either modeled as a *classification* (suitable/unsuitable) or *regression* (suitability measured on an ordinal scale) problem.

While social psychologists concur that *personality* shapes human behavior and can conversely be viewed as a *behavioral construct*, we additionally posit *hireability as a personality construct* in this work. Prior works [6, 9] have put forth this rationale, without rigorously examined the same. Apparent personality trait scores highly correlate with hireability scores, with a linear $R^2 = 0.91$ noted in [6]. The authors also note a categorical HP accuracy of 0.942 from discrete OCEAN personality trait estimates. Likewise, the influence of personality traits on human factors associated with a job profile is noted in [9]. A recent work [25] observes that one's *empathy quotient* (EQ, denoting the drive to empathize) and *systemizing quotient* (SQ, drive to analyze) significantly influence career choices; EQ is associated with the Extraversion and Agreeableness traits [10].

We posit hireability as a function of the *big-five* Openness (O), Conscientiousness (C), Extraversion (E), Agreeableness (A) and Neuroticism (N) personality traits [23]. As in Figure 1, we predict (continuous and discrete) hireability measures from behavioral cues as a two-step process: in the first step, OCEAN personality measures are either derived from apparent (groundtruth) trait ratings, or estimated from textual, audio and visual cues. HP from OCEAN measures is then performed in the second step. Apart from facilitating explanations, we show that HP via this two-step process is more accurate than direct prediction from low-level behavioral cues. Overall, this work makes the following research contributions:

[*] Both authors contributed equally to this research.

ICVGIP '23, December 15–17, 2023, Rupnagar, India

ACM ISBN 979-8-4007-1625-6/23/12…$15.00
https://doi.org/10.1145/3627631.3627658

(1) We expressly and rigorously explore correlations between the OCEAN *personality traits* and *hireability*. While personality may not determine one's profession, links between personality traits and job profiles have been noted [6, 9, 25]. For recruiters, personality assessment would help identify candidates who sync with the job requirements and company culture. While past works [6, 9] have presented 'proof-of-concept' connections between personality and hireability on the *First Impressions Candidate Screening* (FICS) dataset [6], we extensively explore relations among behavioral cues vis-á-vis personality and hireability traits.

(2) With multiple modalities, we show that a two-stage HP (Fig. 1) is superior to end-to-end HP from behavioral cues. This is surprising as end-to-end prediction is less error-prone, thereby fueling the wide-use of deep neural networks. Also, HP via the OCEAN traits would be more *explainable* and *interpretable* than a 'black-box' model fed with high-dimensional behavioral features. Furthermore, these results reveal that accurate personality characterization in-turn enables superior HP.

(3) Different from [6, 9], we present *model-based explanations* to illustrate connections between multiple features and trait prediction via textual and visual analysis. For text, intuitive connections between *word stems* and *trait impressions* are observable. Also, we decompose the visual stream into three parts, capturing the candidate's *facial*, *eye-related* and *bodily* details. Grad-CAM [20] based visualizations show the salience of eye and mouth movements among facial cues, along with head orientation and gestural patterns among bodily cues towards trait estimation. Cumulatively, they enable useful findings; *e.g.*, Impressions of Conscientiousness, which considerably influences hireability, are impacted by both adjectives reflecting *self-discipline* and *achievement* (verbal behavior) and *eye movements* (non-verbal behavior) confirming prior findings [11, 13].

(4) Continuing with (3), eye and bodily cues achieve prediction performance *comparable* to the facial cue, which would encode maximum detail concerning candidate behavior. These results convey that privacy-preserving trait prediction is achievable by concealing the facial appearance.

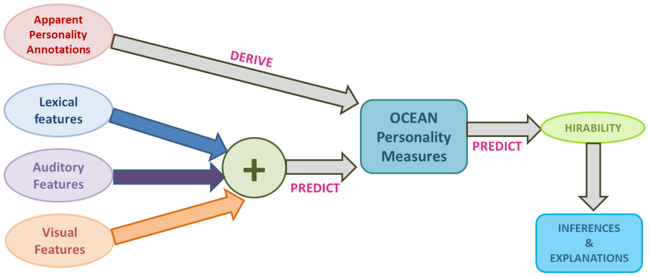

Apparent Personality Annotations
DERIVE
Lexical features
Auditory Features
Visual Features
+
PREDICT
OCEAN Personality Measures
PREDICT
HIRABILITY
INFERENCES & EXPLANATIONS

Figure 1: Study Overview: We posit a significant correlation between suitability for a vocation (termed *hireability*) and personality traits. To this end, OCEAN personality measures are either derived from first-impressions, or predicted from textual, auditory and visual behavioral cues. HP is then achieved from OCEAN measures, and explanations of both hireability and OCEAN trait predictions are attempted.

2 RELATED WORK

We focus on (a) personality trait estimation from behavior and (b) computational hireability prediction in this section.

2.1 Trait prediction from behavioral cues

Social psychologists concur that *personality traits* shape human behavior, and influence many of our life outcomes. Therefore, human-centered intelligent system design has primarily focused on the inverse problem; that of multimodal employing *behavioral cues* to deduce attributes such as the big-five personality traits [11, 22, 24], and related mental health conditions like *depression* [5, 8] and *stress* [7, 14].

Recently, hireability prediction (HP) has been attempted by a number of researchers [6, 9, 17] from multimodal behavioral cues. Fool-proof HP would enable large organizations to employ AHAs and effectively reach out to many applicants on a daily basis. HP algorithms have typically modeled hireability (or *interview* variable 'I') as an adjunct to the OCEAN traits, and predict IOCEAN traits from multimodal behavior.

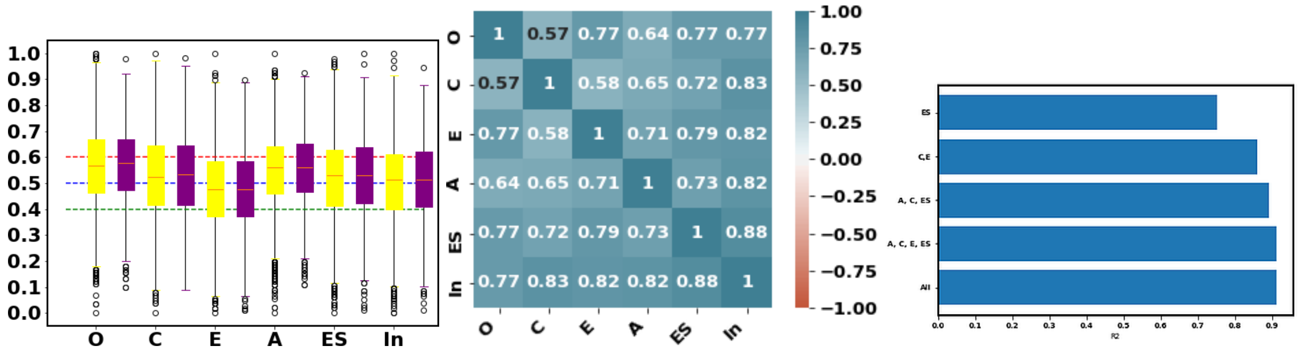
Contrastingly, we posit a strong connect between *hireability* and *personality traits*, as recruiters would typically look for certain traits in candidates reflective of the organization's culture and values. Moreover, recent studies [25] have proposed a connection between the *empathy quotient* psychometric, related to Conscientiousness and Agreeableness, and one's career choices. Inspired by these findings, our work explicitly explores the connection between hireability and the OCEAN traits. Experiments show that HP from OCEAN measures is superior to direct prediction from multimodal behavior.

2.2 Explainable hireability prediction

To ensure transparent recruitment, AHAs need to *justify* their decisions/recommendations, termed *explainability* in machine learning. The few works [6, 9, 17] that have examined HP have focused on quantitatively isolating behavioral IOCEAN correlates. Two recent HP works [6, 9] have loosely explored explainability. Specifically, [6] explains hireability predictions based on personality ratings and categorical OCEAN estimates, while [9] shows typical facial and audio characteristics reflective of apparent traits. Differently, we show (a) how candidates' verbal behavior influences their apparent traits, and (b) what deep neural networks focus on, given candidates' facial and body movements, to explain IOCEAN predictions.

2.3 Inferences from literature survey

Summarizing prior HP works, we note that (1) HP has been attempted from behavioral measures, but not rigorously from personality estimates obtained via behavioral measures; we posit a strong correlation between personality and hireability, and that HP would be superior and explainable if modeled as a function of the OCEAN traits; HP effectiveness from OCEAN measures is confirmed via experiments. (2) Very limited *explainable* evidence pertaining to IOCEAN impressions is available; We perform explanative analyses to show how multimodal verbal and visual cues influence trait impressions, particularly Conscientiousness.



(a) (b) (c)

Figure 2: (a) Boxplots denoting distributions of the OCEAN *personality trait* and the *Interview* (I) measures from the FICS dataset. Train data (8000 videos) are depicted in yellow, and test data (2000 videos) in purple. N trait is denoted as ES. (b) Heatmap depicting correlations among these six attributes. (c) R^2 values obtained for the *best* linear regression model predicting I score with 1–5 personality trait predictors. Best viewed in color and under zoom.

3 OVERVIEW OF THE FICS DATASET

This section is designed to provide readers with an overview of the First Impressions Candidate Screening (FICS) dataset, and serves as a prelude for later sections. Interested readers may refer to [6, 9] for further details.

The FICS video dataset comprises 10000 videos (6000 training, 2000 validation and 2000 testing), and was designed with the objective of developing AHAs to make decisions/recommendations based on multimedia Curriculum Vitae (CVs) [6]. All videos contain labels for *apparent* OCEAN personality traits (reflecting first impressions of a human observer), and a *hireability*/*interview* trait, indicating whether the video candidate should be invited for a job interview. OCEAN and interview (I) scores range within [0,1]. Since Neuroticism (N) is a negative trait, N scores are replaced by Emotional Stability (ES) scores in FICS, and N, ES terms are used interchangeably hereon even if they strictly denote opposite traits.

Fig. 2(a) shows the FICS rating distributions. In all experiments, we trained and validated all our models on the FICS training and validation videos (8000 in total). Roughly similar training and test distributions can be noted for the I, C and E traits from Fig. 2(a). Annotation distributions for all traits are roughly Gaussian, and 70% of A scores fall within one standard deviation from the mean implying a *tight* clustering. The *loosest* clustering is noted for the ES trait, with 67% samples falling within the same range. In terms of inter-quartile range (IQR) denoting the difference between the 75^{th} and 25^{th} percentiles, A has the lowest IQR of 0.18, while C has the highest IQR of 0.22. I score has an IQR of 0.21, and can be seen to be highly correlated with the OCEAN traits from Fig. 2(b). Finally, a linear regression model with OCEAN measures predicting the I score (Fig. 2(c)) shows N as the single-best predictor, and O as the worst. 444 -ve), aggregating 5007 out of the 10K original videos.

In subsequent sections, we predict the I trait as a *continuous* or *categorical* variable from *continuous* OCEAN predictors. We predict both the I and OCEAN scores from multimodal behavioral measures, and show that predicting I from OCEAN estimates is more effective than direct prediction from behavioral cues. To this end, we employ the *estimation accuracy* [9], denoted as Acc = 1 − MAE as the regression/classification metric, where MAE denotes the mean absolute error over the test set. Results outlined in [6] report a maximum Acc of 0.92 for continuous I prediction from audio-visual cues, and an Acc of 0.942 for binarized I prediction from categorical OCEAN estimates.

To examine how visual behaviors affect I and OCEAN impressions, Fig. 3 presents correlations among visual behavioral cues and the IOCEAN traits based on *Openface* [1] outputs. FICS videos are ≈ 15s long, and upon dividing each video into non-overlapping 1s *thin slices* [24], we computed μ, σ statistics for: motion of 68 facial landmarks (μ_{lm}, σ_{lm}), 3D gaze displacement vector ($\mu_{gaze}, \sigma_{gaze}$), eye-gaze pan ($\mu_{gz_x}, \sigma_{gz_x}$), eye-gaze tilt ($\mu_{gz_y}, \sigma_{gz_y}$), head location along (x, y, z) in camera coordinates ($pose_x, pose_y, pose_z$ quantities), head rotation about x, y and z ($Rpose_x, Rpose_y, Rpose_z$ quantities), and the proportion of time for which the candidates' eyes are pointing to the camera/viewer (μ_{ep}, σ_{ep}), over all 1s thin-slices.

Focusing only on significant correlations > 0.1 (highlighted in white) in Fig. 3, we can observe that: (a) consistently high facial movements (μ_{lm}) over thin-slices are positively correlated with the Openness and Extraversion traits (*candidates high on O are expressive, while extraverts typically engage in socially attractive behavior, including facial movements* [15]); (b) high variance in facial motion (σ_{lm}) and eye-gaze pan (σ_{gz_x}) across time-slices correlates negatively with Conscientiousness (individuals high on C exhibit upright gazing behavior, and perseverant impulse control in the presence of social distractors [12]).

Head movements and orientations considerably influence apparent IOCEAN measures. Consistent head motions directed towards and away from the camera (denoted by μ_{pose_z}) indicating a purposeful candidate pitch results in high O,C,N and I scores, while large head motions to the left and right of the camera (denoted by μ_{pose_y}, and suggesting disinterest), correlates negatively with apparent trait measures. Head panning (μ_{Rpose_x}) however is perceived positively by observers, resulting in higher O, C and I scores. Evidently, apparent IOCEAN annotations can be explained by examining candidates' behavior in a fine-grained manner. The next section describes prediction and explanation of apparent IOCEAN measures from multimodal cues.

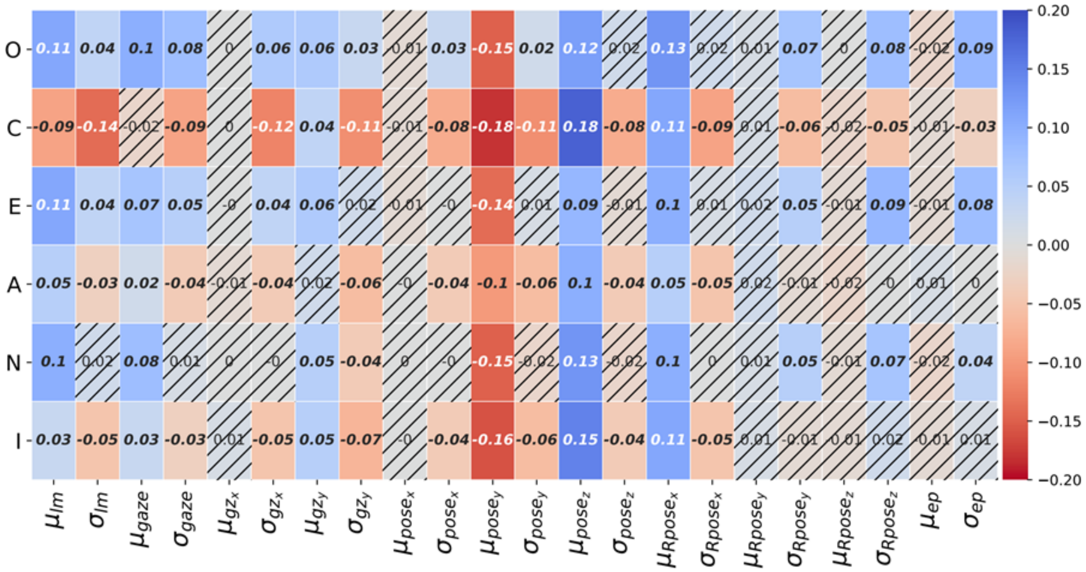


Figure 3: Pearson correlations among visual behavioral cues and IOCEAN annotations for 8000 FICS training videos. Insignificant correlations are crossed out.

4 BEHAVIORAL CUES TO HIREABILITY

This section examines (a) the utility of various language (verbal), auditory and visual cues for HP, (b) compares HP from behavioral cues vis-á-vis the two-step process of personality estimation from behavior, followed by HP from OCEAN estimates, and (c) attempts to explain prediction patterns relating to personality and hireability.

4.1 Verbal (Textual) Cues

As the FICS dataset is accompanied by transcriptions of the candidate videos [6, 9], we examined the impact of candidates' language on their apparent IOCEAN scores.

4.1.1 Experimental Settings. Videos having I/O/C/E/A/N scores ≤ 0.5 considered as negative samples for a given trait, and those with scores > 0.5 were considered as positive samples for classification. For two-step continuous and categorical I score prediction (Table 2), *continuous* OCEAN estimates derived from textual cues were used. The following feature extraction and regressor/classifier frameworks were examined.

Bag of Words (BoW) feature extraction: As in [9], we adopted the BoW approach for text analyses. From video transcripts, stopwords were removed and we used adjectives, adverbs, verbs and nouns to construct our vocabulary. The top 5000 most frequently appearing words overall were selected as features; each transcript is thus denoted by a 5000-D vector, which was input to the following algorithms.

Regressor/classifier frameworks: For regression, we employed the random forest (RF) and Support vector Regressor (SVR), while for classification, we used the (a) Naive Bayes (NB) classifier provided by NLTK (https://www.nltk.org/), (b) Binomial Naive Bayes (B-NB), (c) Logistic Regression (LR), (d) Support Vector Classifier (SVC) and (e) AWD-LSTM, the stochastic gradient descent-based long short-term memory pipeline provided by FastAI (https://www.fast.ai/).

4.1.2 Quantitative and Qualitative Results. Table 1 presents regression (R) and classification (C) on the IOCEAN traits. Table 2 presents continuous/categorical I score prediction from regression-based OCEAN estimates in Table 1. Furthermore, we found the top 10 *most informative* word stems for each trait via the NB classifier (Table 3). Informative word stems were identified as follows. We computed the relative likelihood of candidate selection (S) vs rejection (R) given *stem* via importance weights (IW) as $\text{IW}_{stem} = \Pr(S \mid stem)/\Pr(R \mid stem)$. Therefore, $\text{IW}_{stem} = 10$ implies that the selection likelihood of a candidate using the word *stem* is 10 times *higher* than one without the *stem*; In short, *stems with +ve IWs positively impact trait impressions, while stems with -ve IWs elicit negative trait impressions.* +ve and -ve stems for the IOCEAN traits are respectively coded in green and red in Table 3.

4.1.3 Discussion. From Tables 1,2,3, we make the following remarks.

(1) Continuous IOCEAN values (denoted via R) are better estimated by all models than categorical (C) prediction, as per the Acc values in Table 1. SVR performs better than the RF regressor, achieving Acc$\geq$ 0.9 for the O and A traits. This performance is superior to results achieved via language models in [9].
(2) One can also note from Table 1 that continuous IOCEAN score prediction from textual features (max Acc = 0.904) is more effective than categorical IOCEAN label prediction (max Acc = 0.678). Among classifiers, SVC is most effective achieving a mean Acc of 0.64 across the IOCEAN dimensions; O and A are the two traits best predicted best by all classifiers. These results are attributable to the fact that traits other

Table 1: Quantitative IOCEAN prediction from textual cues.

Model	I	O	C	E	A	N
RF (R)	0.886	0.892	0.881	0.880	0.896	0.883
SVR (R)	0.892	0.900	0.888	0.882	**0.904**	0.896
NB (C)	0.589	0.643	0.638	0.586	0.614	0.594
B-NB (C)	0.599	0.655	0.582	0.591	0.633	0.588
LC (C)	0.594	0.641	0.595	0.579	0.617	0.587
SVC (C)	0.639	**0.678**	0.613	0.599	0.671	0.627
AWD-LSTM (C)	0.577	0.602	0.590	0.583	0.605	0.582

Table 2: HP from continuous OCEAN measures (via textual cues).

Model	RF	SVR	SVC
Regression	0.903	0.903	-
Classification	-	0.646	**0.649**

Table 3: Exemplar +ve (green) and -ve (red) word stems for the IOCEAN traits. IWs specified in brackets.

I	*discuss* (8.6)	*assist* (-7.2)	*self* (5.8)	*boot* (-5.7)
O	*assist* (-12.8)	*boot* (-10.1)	*danger* (-8.8)	*fashion* (7.7)
C	*self* (8.9)	*discuss* (7.3)	*achiev* (5.9)	*allow* (5.5)
E	*explor* (8.7)	*shadow* (-7.8)	*lucki* (7.2)	*aspir* (5.4)
A	*die* (-7.2)	*repli* (6.6)	*graduat* (6.0)	*maintain* (-5.8)
ES (N)	*societi* (-7.9)	*discuss* (7.7)	*wave* (-6.3)	*great* (5.6)

than O and A are densely clustered about 0.5 (Fig. 2), and dichotomization about 0.5 produces many *gray* samples.

(3) Comparing Table 1 with Table 2, we note that continuous I score prediction from OCEAN estimates (Acc of 0.903 in Table 2) is better than direct estimation from textual features (max Acc of 0.892 in Table 1). Likewise, superior binary I classification is achieved from OCEAN predictors (max Acc of 0.649), as compared to direct classification employing text features (max Acc of 0.639). These results convey that a two-step process for I score prediction is optimal.

(4) Also, ***intuitive connections*** between word stems and traits are noted via IWs. Stems such as *discuss* and *self* are seen positively in the context of hireability, while *assist* is perceived negatively. Therefore, individuals effusing independence and keen to collaborate appear favorable from a hireability viewpoint. Stems like *self* and *achiev*, indicating a responsible and goal-driven nature characterize Conscientiousness, while *discuss* positively correlates with emotional stability. Negative stems such as *die* convey low apparent A; the *fashion* stem conveys high creativity, while words such as *assist* and *danger* appear to convey a conservative/traditional mindset.

Table 4: Description of extracted audio features.

MFCCs	Form a representation where frequency bands are not linear but distributed on the mel-scale
Energy	Squared-sum of signal values, normalized by the frame length
ZCR	Zero crossing rate of the signal within a particular frame
Tempo	Beats per minute
Spectral flatness	Measure to quantify *noise-like* trait of a sound spectrum
Spectral bandwidth	p'th-order spectral bandwidth, default $p = 2$
Spectral roll-off	Frequency below which 90% spectrum is concentrated
Spectral contrast	For each sub-band, compare mean energy of top quantile with mean of bottom quantile.
Tonnetz	Tonal centroid features

4.2 Auditory cues

4.2.1 Feature extraction. For predicting IOCEAN traits from audio, we extracted low-level speech signal statistics from the *Librosa* library (https://librosa.github.io/librosa/feature.html), and audio spectrograms. *Librosa* features were fed to a random forest (RF), while speech sprectrograms were fed to a VGG11 (CNN) for regression/classification as in Table 5. A total of 56 audio statistics including μ, σ for 20 MFCC coefficients (Table 4) were employed for analysis.

4.2.2 Experimental Settings. For IOCEAN estimation, we considered the IOCEAN traits as both continuous and categorical; regression and classification results are coded as (R) and (C) respectively in Table 5. As a second step, we predicted continuous/categorical I scores from continuous/discrete OCEAN estimates (Table 6). We also adopted the *thin-slice* approach (as in Sec. 3) for audio analysis, aggregating 1s *Librosa* statistics over 2–15 second time-windows to predict continuous IOCEAN measures (Fig. 4). Results in Table 5,6 correspond to 15s time windows (equal to the length of FICS videos).

Table 5: Audio performance for IOCEAN estimation.

Model	I	O	C	E	A	N
RF (R)	**0.9043**	0.9036	0.0.8986	0.9000	0.9054	0.9001
CNN (R)	0.8967	0.8966	0.8905	0.8930	0.9006	0.8944
RF (C)	0.7140	0.6950	**0.7150**	0.6950	0.6545	0.7125
CNN (C)	0.6725	0.6700	0.6555	0.6590	0.6370	0.6860

Table 6: HP from OCEAN measures (audio cues). Labels R and C denote continuous/categorical OCEAN estimates.

Regression		Classification			
RF (R)	CNN (R)	RF (R)	CNN (R)	RF (C)	CNN (C)
0.9047	0.8899	0.7040	0.6575	0.7015	**0.7865**

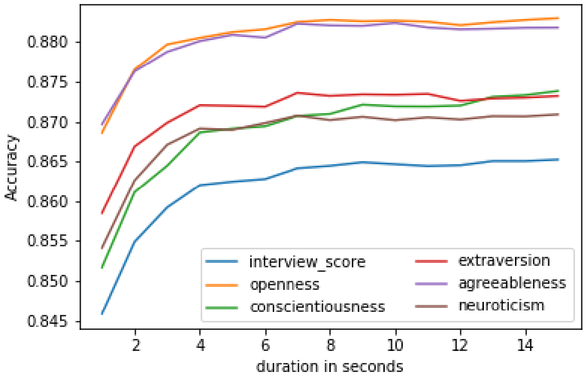


Figure 4: IOCEAN prediction from *Librosa* features with varying time windows.

4.2.3 Results and Discussion. We make the following remarks from our experimental results. (1) As with text analysis, continuous IOCEAN prediction is better achieved (max Acc = 0.8916) than discrete (max Acc = 0.8116). (2) Consistent with text-based results, better prediction of I scores is achieved from continuous OCEAN estimates (max Acc = 0.8946), than from audio features (max Acc = 0.8799). (3) The time-window varying experiment was designed to verify if 15s of audio data is indeed necessary for accurate IOCEAN prediction. From Fig. 4, we note that the Acc results saturate beyond 6s, reflecting that reliable trait estimation is achievable upon observing only *tiny* behavioral episodes, and conveying that 15s windows is redundant for audio-based trait estimation. Overall, the O and A traits are best reflected by audio features, while Interview scores are not well predicted via *Librosa* statistics.

4.3 Visual Analysis

Non-verbal behavior cues, especially visual, have been extensively employed for human-centered applications earlier [5, 7, 22, 24]. This is due to the fact that visual behaviors such as gazing, facial emotions and movements, and body movements convey a significant amount of informative and communicative cues during social interactions. Especially during interview sessions, visual behaviors can convey a lot of information (*is the candidate calm or emotional when facing a tough situation?*) to the interviewer.

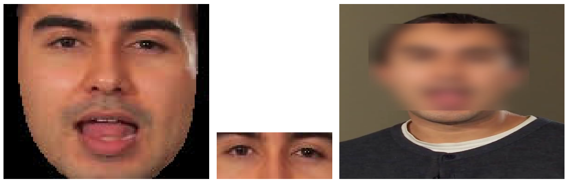


Figure 5: Inputs to the visual model include the cropped face image (left), cropped eye region (center) and face-blurred portrait to examine the influence of holistic body movements for trait prediction.

Given the critical contribution of visual behavior to IOCEAN prediction, we opted to examine *multiple* visual cues different from prior HP works [6, 9, 17]. Instead of examining only facial cues for trait prediction, we also proceeded to examine the *eye* and the *body movements*; we therefore additionally input an eye-crop and a body-crop with the face blurred (Fig. 5) to the prediction frameworks, to evaluate the contribution of eye and body movements towards IOCEAN prediction. The face and eye-crops are obtained via *Openface* [1], while the face-blurred body-crop is obtained by smoothing the facial region in the video frame using a Gaussian filter, so that the facial details are not apparent to the observer.

4.3.1 Experimental settings. We considered the following prediction models in our experiments.

2D-CNN: A 19 layered VGG model, which processes 2D frame information was used. The VGG output layer was removed, and two hidden fully-connected layers with 512 and 64 neurons respectively were added along with output layer involving 6 neurons (one neuron each for the IOCEAN traits). Mean squared error (MSE) for regression, and binary cross-entropy (BCE) loss for classification were used during training on a single, representative frame from the video sequence, with learning rate of 1e-4 and a batch size of 64.

3D-CNN: An 18 layered ResNet-3D model (https://arxiv.org/abs/1711.11248), with pre-trained weights for human activity recognition, was used. The 3D-CNN model took inputs from 16 uniformly spaced visual frames, sampled at $t = 0, 1, \ldots, 15$ seconds into the video. The ResNet-3D output layer was removed, and two hidden fully-connected layers with 128 and 32 neurons respectively were added instead, along with final output layer of 6 neurons. The 16 stacked frames are re-sized to $112x112$ prior to input. Mean squared error (MSE) Loss was used during training (3D-CNN was employed only for regression), with learning rate 1e-4 and batch size 32.

LRCN: which denotes a Long-term Recurrent Convolutional Neural network (https://arxiv.org/abs/1411.4389) with a pre-trained ResNet-50 encoder and a single-layer LSTM decoder. This model takes 40 uniformly-spaced video frames as input; the encoder CNN learns 512-D features for each frame, which are fed to the LSTM decoder across different time frames. The 512-D LSTM output is fed into a linear layer of size 256, which is then connected to the final layer composed of 6 neurons. L1-loss was used for model training, with learning rate for the pre-trained ResNet set to 1e-6, and varying between 1e-4 to 1e-5 for other layers. The Adam optimizer was used to train the LRCN.

Table 7: IOCEAN Regression from visual cues: 2D, 3D and LRC refer to 2D-CNN, 3D-CNN and LRCNN. F, E and B denote facial, eye and body cues.

	2D(F)	2D(E)	2D(B)	3D(F)	3D(E)	3D(B)	LR(F)	LR(E)	LR(B)
I	0.9111	0.9009	0.9159	**0.9206**	0.9106	0.9175	0.9140	0.9034	0.9111
O	0.9056	0.8997	0.9097	**0.9136**	0.9055	0.9124	0.9031	0.9008	0.9074
C	0.9084	0.8993	0.9159	**0.9207**	0.9113	0.9183	0.9027	0.9001	0.9126
E	0.9076	0.8979	0.9110	**0.9212**	0.9095	0.9133	0.9078	0.9006	0.9068
A	0.9057	0.8998	0.9111	**0.9134**	0.9089	0.9106	0.9102	0.9027	0.9078
N	0.9026	0.8948	0.9071	**0.9124**	0.9030	0.9083	0.9022	0.8965	0.9034

Table 8: IOCEAN Classification from visual cues: 2D, 3D and LRC refer to 2D-CNN, 3D-CNN and LRCNN respectively. Codes F, E and B denote facial, eye and body cues.

	2DC(F)	2DC(E)	2DC(B)	3DC(F)	3DC(E)	3DC(B)	LRC(F)	LRC(E)	LRC(B)
I	0.7130	0.6648	0.6855	**0.7600**	0.6933	0.7460	0.7365	0.6765	0.7200
O	0.6965	0.6563	0.6725	0.7150	0.6853	0.7265	**0.7360**	0.6690	0.7073
C	0.7050	0.6693	0.6815	**0.7730**	0.6978	0.7645	0.7435	0.6848	0.7120
E	0.7120	0.6643	0.6780	0.7380	0.6848	**0.7410**	0.7315	0.6848	0.6940
A	0.6560	0.6158	0.6495	0.6920	0.6508	0.6780	**0.7125**	0.6305	0.6660
N	0.7165	0.6653	0.6785	**0.7425**	0.6823	0.7390	0.7230	0.6800	0.6823

Table 9: HP from *continuous* OCEAN estimates. 2D, 3D and LR refer to 2D-CNN, 3D-CNN and LRCNN. F, E and B codes in brackets stand for facial, eye and body cues. R/C codes denote continuous/categorical HP.

2D(FR)	2D(ER)	2D(BR)	3D(FR)	3D(ER)	3D(BR)	LR(FR)	LR(ER)	LR(BR)
0.9120	0.9021	0.9167	**0.9220**	0.9106	0.9184	0.9167	0.9032	0.9111
2D(FC)	**2D(EC)**	**2D(BC)**	**3D(FC)**	**3D(EC)**	**3D(BC)**	**LR(FC)**	**LR(EC)**	**LR(BC)**
0.7355	0.6758	0.7415	**0.7745**	0.7099	0.7600	0.7400	0.7064	0.7611

4.3.2 Results & Discussion. Tables 7, 8 and 9 present trait predictions from the multiple visual cues. From Table 7, which estimates continuous IOCEAN values from the face, eye and body cues, we make the following remarks: (1) In terms of the general predictive power, the 3D-CNN is more potent than the 2D-CNN and LRCNN frameworks. Acc values ≥ 0.9 are often observed with the 3D-CNN, while the 2D-CNN and LRCNN perform slightly inferiorly. (2) An interesting finding is that the eye and body-cues achieve performance *comparable* to the face cue. This is particularly important as it opens up the possibility of AHAs being able to examine video CVs and make reasonable trait-related decisions *while honoring the*

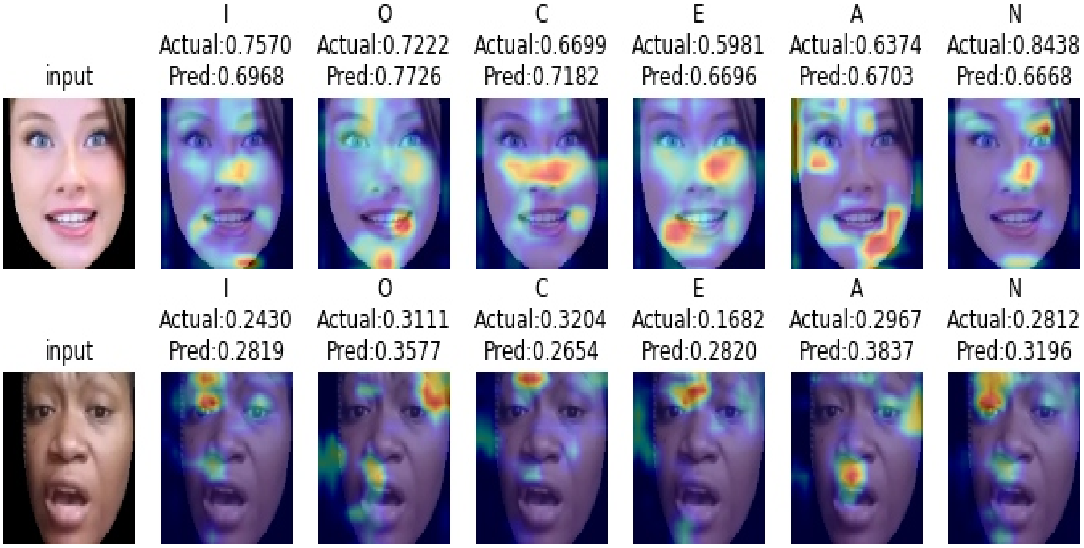


Figure 6: Exemplar grad-cam outputs for a person eliciting *high* trait scores (top) and *low* trait scores (bottom). Eyes are the primary cue for eliciting apparent Conscientiousness impressions, while other traits are influenced by holistic facial structure and facial emotions. Best-viewed in color.

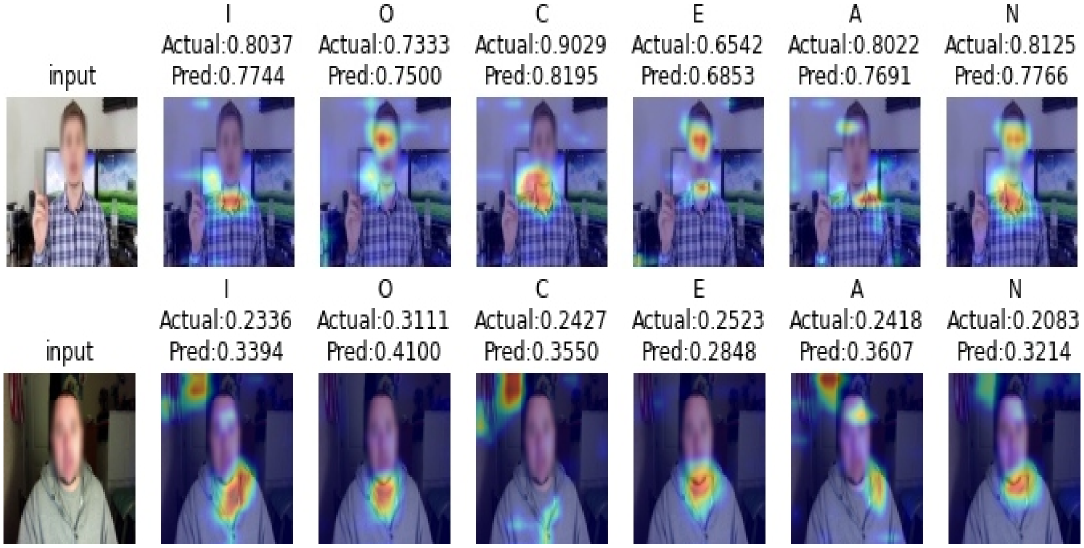


Figure 7: Exemplar grad-cam outputs on blurred face portraits for a person eliciting *high* trait scores (top) and *low* trait scores (bottom). Attention maps indicate a focus on the neck region, which determines the relative orientation between the face and body, hand gestures and clothing. Best-viewed in color.

candidate's privacy (processing only a mid-to-low resolution image of the eye, or blurring the face will render the facial information unusable as a biometric). (4) The face cue is nevertheless critical, and produces the best prediction for the Interview trait. (5) Among OCEAN traits, C and A are the two best-predicted traits from visual cues.

Focusing on IOCEAN classification results in Table 8, in line with the text and audio-based results, considerably lower Acc values than regression are noted for classification. Interestingly, body cues produce the best categorical IOCEAN estimates, and achieve considerably better performance than face or eye cues. This results indicates that perhaps, *a fine-grained visual examination of the candidate's behavior may not be necessary to make a coarse-grained decision (i.e., suitable or unsuitable) regarding the candidate's hireability.* A distant examination could still be adequate. Among IOCEAN traits, N is predicted best based on body cues by the 2D-CNN, which is revealing as the N trait is associated with anxiety, which may manifest via body-fidgeting, *etc.*

Examining Table 9 which presents continuous/categorical HP from OCEAN estimates, we again note that Acc ≥ 0.9 is achieved for all conditions (second table row), except with the 2D-CNN employing eye information. The best prediction of categorical I labels (Acc = 0.87) is achieved when continuous OCEAN scores are estimated employing facial information; this implies that ***reasonable coarse-grained hireability decisions are possible even when accurate OCEAN estimates are available to the AHA in lieu of a multimedia CV***.

4.3.3 Explaining visual Predictions. While the above inferences may be logically derived from experimental results, we explored if any explanations of the visual predictions are possible. Prior works [6, 9] show some visual correlates of the IOCEAN traits without *explicitly* showing where their predictive models are looking at. Differently, we employed the Grad-CAM algorithm [20] to *highlight* image regions deemed important for a trait prediction. Using Grad-CAM, gradients of the IOCEAN output neurons are used to get a weighted-sum of the convolutional layer output maps, termed *attention maps* depicting where the network *sees* to accurately predict the trait. We generated activation maps for the IOCEAN traits highlighting important facial and body cues (Figures 6, 7).

Fig. 6 shows Grad-CAM outputs for a *high* and a *low* trait exemplar. One can note that the attention maps relate to the eye and the mouth regions for the IOCEAN traits, which are likely to be of interest to a human interviewer as well. Conscientiousness is one (possible) exception where attention is more localized to the eyes. Conscientiousness is associated with sincerity and uprightness, and is traditionally gauged from eye-movement cues [11]. Conversely, when the face is blurred so as to make the facial cues indecipherable (Fig. 7), the activation maps are focused around the neck region, hand movements and clothing. When the face is represented as a blob, the neck region becomes important as it determines the relative orientation between the face and body. ***These visual explanations cumulatively convey the importance of eye and mouth movements, hand gestures and attire for HP.***

5 DISCUSSION & CONCLUSION

At the outset, the objectives of this work were two-fold: (1) to explicitly and rigorously explore the correlations between *hireability* and the OCEAN *personality traits*, given that this dependence has been exploited earlier in a limited way [6, 9], and (2) to provide *explanations* supporting IOCEAN predictions made by the multimodal behavioral models. Based on the experimental results, we conclude that this work has substantially achieved both objectives.

With respect to (1), we note that continuous/categorical HP from OCEAN estimates, which are in-turn obtained from audio, visual and verbal behaviors, is more effective than directly predicting from behavioral measures. While this may seem surprising, we believe that this result is only an implication of designing a *simple* HP model with only the OCEAN trait predictors, rather than a 'black-box' model with high-dimensional inputs but limited interpretability.

Regarding (2), we note that all considered modalities and features provide some explanations towards IOCEAN prediction. With respect to *text*, we found that IWs of word stems are highly informative; *e.g.*, use of the word *dead* negatively impacts hiring impressions, and conveys anxiety (indicator of Neuroticism). The words *hobbi* and *fashion* convey a high level of Extraversion. Apparent Conscientiousness is negatively impacted by cuss words, but positively by words relating to well-being. While audio-related explanations are not explicitly presented, we note from Figure 4 that IOCEAN predictions saturate beyond 6s time-windows, implying that *tiny* behavioral episodes suffice for reliable trait prediction.

Visual cues are also highly informative, as confirmed by both quantitative and qualitative results. Quantitative results show that the *eye* and *body* cues achieve IOCEAN prediction comparable to *face* cues. This is a useful result, as processing facial information incapable of revealing identity would assuage candidates' privacy concerns. That body cues can achieve high accuracy on categorical IOCEAN prediction implies that fine-grained behavioral analytics may not be necessary for making coarse-grained decisions. Also, Table 9 conveys that coarse hiring decisions are possible solely based on a candidate's OCEAN estimates. Grad-CAM visualizations show the influence of *eye* and *mouth movements*, *hand movements* and *attire* on hireability.

A limitation of this study is that our experiments are only performed on the FICS dataset. Future work will focus on validating, extending and generalizing current results via experimentation on multiple datasets.

REFERENCES

[1] Tadas Baltrusaitis, Amir Zadeh, Yao Lim, and Louis-Philippe Morency. 2018. OpenFace 2.0: Facial Behavior Analysis Toolkit. 59–66. https://doi.org/10.1109/FG.2018.00019

[2] Ligia Batrinca, Bruno Lepri, Nadia Mana, and Fabio Pianesi. 2012. Multimodal Recognition of Personality Traits in Human-Computer Collaborative Tasks. In *International Conference on Multimodal Interaction* (Santa Monica, California, USA) *(ICMI '12)*. Association for Computing Machinery, New York, NY, USA, 39–46. https://doi.org/10.1145/2388676.2388687

[3] Maneesh Bilalpur, Mohan Kankanhalli, Stefan Winkler, and Ramanathan Subramanian. 2018. EEG-Based Evaluation of Cognitive Workload Induced by Acoustic Parameters for Data Sonification. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction* (Boulder, CO, USA) *(ICMI '18)*. Association for Computing Machinery, New York, NY, USA, 315–323. https://doi.org/10.1145/3242969.3243016

[4] Maneesh Bilalpur, Seyed Mostafa Kia, Manisha Chawla, Tat-Seng Chua, and Ramanathan Subramanian. 2017. Gender and Emotion Recognition with Implicit User Signals. In *International Conference on Multimodal Interaction* (Glasgow, UK) *(ICMI '17)*. Association for Computing Machinery, New York, NY, USA, 379–387. https://doi.org/10.1145/3136755.3136790

[5] Nicholas Cummins, Julien Epps, Michael Breakspear, and Roland Goecke. 2011. An Investigation of Depressed Speech Detection: Features and Normalization. *Proc. Interspeech*, 2997–3000.

[6] Jair Hugo Escalante, Meysam Madadi, Stephane Ayache, Evelyne Viegas, Furkan Gurpinar, Sukma Achmadnoer Wicaksana, Cynthia Liem, A. J. Van Marcel Gerven, Van Rob Lier, Heysem Kaya, Ali Albert Salah, Sergio Escalera, Yagmur Gucluturk, Umut Guclu, Xavier Baro, Isabelle Guyon, and C. S. Julio Jacques. 2020. Modeling, Recognizing, and Explaining Apparent Personality from Videos. *IEEE Transactions on Affective Computing* (2020), 1–1.

[7] Ailbhe N. Finnerty, Skanda Muralidhar, Laurent Son Nguyen, Fabio Pianesi, and Daniel Gatica-Perez. 2016. Stressful First Impressions in Job Interviews. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction* (Tokyo, Japan) *(ICMI '16)*. Association for Computing Machinery, New York, NY, USA, 325–332. https://doi.org/10.1145/2993148.2993198

[8] Monika Gahalawat, Raul Fernandez Rojas, Tanaya Guha, Ramanathan Subramanian, and Roland Goecke. 2023. Explainable Depression Detection via Head Motion Patterns. In *Int'l Conference on Multimodal Interaction* (Paris, France). 261–270. https://doi.org/10.1145/3577190.3614130

[9] Yagmur Gucluturk, Umut Guclu, Xavier Baro, Hugo Jair Escalante, Isabelle Guyon, Sergio Escalera, Marcel A. J. van Gerven, and Rob van Lier. 2018. Multimodal First Impression Analysis with Deep Residual Networks. *IEEE Trans. Affect. Comput.* 9, 3 (July 2018), 316–329. https://doi.org/10.1109/TAFFC.2017.2751469

[10] Brian W. Haas, Michael Brook, Laura Remillard, Alexandra Ishak, Ian W. Anderson, and Megan M. Filkowski. 2015. I Know How You Feel: The Warm-Altruistic Personality Profile and the Empathic Brain. *PLOS ONE* 10, 3 (03 2015), 1–15. https://doi.org/10.1371/journal.pone.0120639

[11] Sabrina Hoppe, Tobias Loetscher, Stephanie A Morey, and Andreas Bulling. 2018. Eye Movements During Everyday Behavior Predict Personality Traits. *Frontiers*

in human neuroscience 12, 105 (2018).

[12] Sabrina Hoppe, Tobias Loetscher, Stephanie A Morey, and Andreas Bulling. 2018. Eye movements during everyday behavior predict personality traits. *Frontiers in human neuroscience* (2018), 105.

[13] Timothy Jay and Kristin Janschewitz. [n.d.]. The Science of Swearing. https://www.psychologicalscience.org/observer/the-science-of-swearing.

[14] Radhika Kuttala, Ramanathan Subramanian, and Venkata Ramana Murthy Oruganti. 2023. Multimodal Hierarchical CNN Feature Fusion for Stress Detection. *IEEE Access* 11 (2023), 6867–6878. https://doi.org/10.1109/ACCESS.2023.3237545

[15] Bruno Lepri, Ramanathan Subramanian, Kyriaki Kalimeri, Jacopo Staiano, Fabio Pianesi, and Nicu Sebe. 2012. Connecting meeting behavior with extraversion—A systematic study. *IEEE Transactions on Affective Computing* 3, 4 (2012), 443–455.

[16] Kristiyan Lukanov, Horia A. Maior, and Max L. Wilson. 2016. Using FNIRS in Usability Testing: Understanding the Effect of Web Form Layout on Mental Workload. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 4011–4016. https://doi.org/10.1145/2858036.2858236

[17] Iftekhar Naim, Md. Iftekhar Tanveer, Daniel Gildea, and Mohammed (Ehsan) Hoque. 2018. Automated Analysis and Prediction of Job Interview Performance. *IEEE Trans. Affect. Comput.* 9, 2 (2018), 191–204. https://doi.org/10.1109/TAFFC.2016.2614299

[18] Viral Parekh, Pin Sym Foong, Shengdong Zhao, and Ramanathan Subramanian. 2018. AVEID: Automatic Video System for Measuring Engagement In Dementia. In *23rd International Conference on Intelligent User Interfaces* (Tokyo, Japan) *(IUI '18)*. Association for Computing Machinery, New York, NY, USA, 409–413. https://doi.org/10.1145/3172944.3173010

[19] Eric Rosenbaum. [n.d.]. IBM artificial intelligence can predict with 95% accuracy which workers are about to quit their jobs. https://www.cnbc.com/2019/04/03/ibm-ai-can-predict-with-95-percent-accuracy-which-employees-will-quit.html.

[20] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*. 618–626.

[21] A. Shukla, S. S. Gullapuram, H. Katti, M. Kankanhalli, S. Winkler, and R. Subramanian. 2020. Recognition of Advertisement Emotions with Application to Computational Advertising. *IEEE Transactions on Affective Computing* (2020), 1–1.

[22] Ramanathan Subramanian, Jacopo Staiano, Kyriaki Kalimeri, Nicu Sebe, and Fabio Pianesi. 2010. Putting the Pieces Together: Multimodal Analysis of Social Attention in Meetings. In *International Conference on Multimedia* (Firenze, Italy) *(MM '10)*. Association for Computing Machinery, New York, NY, USA, 659–662. https://doi.org/10.1145/1873951.1874045

[23] R. Subramanian, J. Wache, M. K. Abadi, R. L. Vieriu, S. Winkler, and N. Sebe. 2018. ASCERTAIN: Emotion and Personality Recognition Using Commercial Sensors. *IEEE Transactions on Affective Computing* 9, 2 (2018), 147–160. https://doi.org/10.1109/TAFFC.2016.2625250

[24] Ramanathan Subramanian, Yan Yan, Jacopo Staiano, Oswald Lanz, and Nicu Sebe. 2013. On the Relationship between Head Pose, Social Attention and Personality Prediction for Unstructured and Dynamic Group Interactions. In *International Conference on Multimodal Interaction*. Association for Computing Machinery, New York, NY, USA, 3–10. https://doi.org/10.1145/2522848.2522862

[25] Alessandro Vinciarelli, Walter Riviera, Francesca Dalmasso, Stefan Raue, and Chamila Abeyratna. 2019. What Do Prospective Students Want? An Observational Study of Preferences About Subject of Study in Higher Education. In *Innovations in Big Data Mining and Embedded Knowledge*, Anna Esposito, Antonietta Maria Esposito, and Lakhmi C. Jain (Eds.). Intelligent Systems Reference Library, Vol. 159. Springer, 83–97. https://doi.org/10.1007/978-3-030-15939-9_5