Understanding Applicants' Reactions to Asynchronous Video Interviews Through Self-reports and Nonverbal Cues

Skanda Muralidhar
Idiap Research Institute
Switzerland

Emmanuelle Patricia Kleinlogel
University of Lausanne
Switzerland

Eric Mayor
University of Basel
Switzerland

Adrian Bangerter
University of Neuchatel
Switzerland

Marianne Schmid Mast
University of Lausanne
Switzerland

Daniel Gatica-Perez
Idiap Research Institute & EPFL
Switzerland

ABSTRACT

Asynchronous video interviews (AVIs) are increasingly used by organizations in their hiring process. In this mode of interviewing, the applicants are asked to record their responses to predefined interview questions using a webcam via an online platform. AVIs have increased usage due to employers' perceived benefits in terms of costs and scale. However, little research has been conducted regarding applicants' reactions to these new interview methods. In this work, we investigate applicants' reactions to an AVI platform using self-reported measures previously validated in psychology literature. We also investigate the connections of these measures with nonverbal behavior displayed during the interviews. We find that participants who found the platform creepy and had concerns about privacy reported lower interview performance compared to participants who did not have such concerns. We also observe weak correlations between nonverbal cues displayed and these self-reported measures. Finally, inference experiments achieve overall low-performance w.r.t. to explaining applicants' reactions. Overall, our results reveal that participants who are not at ease with AVIs (i.e., high creepy ambiguity score) might be unfairly penalized. This has implications for improved hiring practices using AVIs.

CCS CONCEPTS

• **Applied computing** → **Psychology**; • **Human-centered computing** → *Empirical studies in HCI*.

KEYWORDS

Social computing; nonverbal behavior; self-reported measures; multimodal interaction; video interviews

1 INTRODUCTION

Asynchronous video interviews (AVIs) are gaining attention as a tool in the employee recruitment process. This process is helping recruiters assess and short-listing potential candidates with lower costs. While the feasibility of developing systems for automatic behavioral assessment using nonverbal and verbal cues displayed by applicants during AVIs has been investigated in the literature, there has been little work in understanding applicants' reactions towards this new interview method. In this work, we aim to bridge this gap by collecting a novel dataset, analyze self-reported questionnaires of applicants' reactions and its connections to automatically extracted nonverbal behavior in simulated AVIs.

AVIs are being offered as a service by several companies. The overall procedure is usually comparable and consists of applicants connecting to a website and answering a sequence of questions predefined by the recruiter while recording a video of themselves using a camera (like those in a laptop, smartphone, or tablet). These videos are then either watched and rated by recruiters connecting to the same platform, or use automated methods to short-list applicants. These applicants could then be called for further interview processes including face-to-face interviews. Such platforms claim to provide several advantages: they allow the applicants to take the video interviews at time and place of their choice while allowing organizations to screen applicants faster and at reduced costs.

Existing work in computing so far has focused on automatic analysis of applicants' nonverbal behavior either for behavior assessment [12, 13, 21, 22, 32, 38] or feedback [1, 24, 37]. It has previously been shown that participants are perceived differently in AVIs versus face-to-face settings along several dimensions including communication skills, fluency, and enthusiasm [45]. However, there is little research towards understanding applicants' reactions to such interview methods or capturing these reactions automatically. This understanding is important as some research in organizational psychology has reported a negative impact of such technologically mediated interview methods in terms of applicants' reaction and interviewee performance ratings [11, 29]. There is also increasing concern on the fairness of these platforms as currently used in commercial settings [44, 48].

Applicants' reactions are an important construct in personnel selection research. This is because literature has shown that perceptions about the interview process (including the forms of interview used) influences applicant reactions during hiring (like company attractiveness, and job acceptance) and after hiring (like on-the-job performance, attitudes, and reapplication intentions) [6, 7]. In this work, we use the definition proposed in [47] which denotes applicants' reactions as "attitudes, affect or cognitions an individual might have about the hiring process." Under this broad definition, we are interested in applicants' perceptions of procedural justice rules, global fairness, privacy concerns, creepiness, job-relatedness, and self-perceptions (i.e. interview performance). Specifically, we pose the following research questions:

RQ1: How do young applicants' react to AVIs? Does this influence their self-reported performance?

RQ2: What is the underlying structure of applicants' reactions?

RQ3: Are there any connections between applicants' reactions and their displayed nonverbal behavior?

The contributions of this work are the following. First, we describe the design of a study to integratively analyze applicants' reactions to AVIs. Second, we collected a novel dataset of AVIs with 221 participants with 4 questions per participant eliciting different behavioral responses. Third, we analyzed the self-reported measures to understand applicants' reactions to this mode of employee selection. We then performed a Principal Component Analysis (PCA) on the various measures to capture the underlying basic statistical structure. Fourth, we analyzed the relationship between automatically extracted nonverbal cues and some self-reported measures in a correlation analysis and in an inference task. Our results reveal that participants who are not at ease with this interview method (i.e., high creepy ambiguity score) might be unfairly penalized when participating in AVIs. This has implications for improved hiring practices using AVIs.

We believe this work constitutes a step forward in understanding the impact of AVIs on applicants. Specifically, our work shows that applicants' reactions are an import aspect to consider and may potentially influence applicant's performance during AVIs.

2 LITERATURE REVIEW

Based on the type of interaction, job interviews can be categorized into *face-to-face interviews*, *online videos*, *asynchronous video interviews*. All the categories of interviews have been studied in the computing literature from two main perspectives: automatic behavioral analytics; and feedback and training systems.

Face-to-face interviews are the traditional form of hiring where interviewer and interviewee sit facing each other. This setting has been widely studied from the perspective of behaviour analytics [12, 32, 37, 38]. Nguyen et al. investigated the performance of automatically extracted nonverbal features in inferring perceived hirability [38]. The authors, using a dataset of 62 real interview videos, reported a performance of $R^2 = 0.36$. This work was extended by Naim et al., who using a dataset of 138 simulated interview videos, investigated verbal and nonverbal behavior [37]. The authors reported that participants who spoke fluently with less filler words, used more unique words, and smiled more were perceived to be more hirable. Muralidhar et al. designed and developed a behavioral

training procedure to help hospitality students improve their first impressions [32]. Using this living lab, they collected 169 simulated interviews and reported an inference performance on perceived hirability of $R^2 = 0.32$ using nonverbal behavioral cues. From the standpoint of feedback and training, Hoque et al. [24] developed My Automated Conversation Coach (MACH). MACH is a virtual agent that helps trainees' improve their real interview performance by automatically sensing facial expressions, speech content, and prosody, and respond with appropriate verbal and nonverbal cues in real time.

Online videos are short videos in which job applicants present themselves and post them on social media to communicate with potential employers[23]. Due to wide-spread acceptance of social media (like YouTube) and hence online interviews, this has allowed investigation of work-related social constructs at large scale. The widespread popularity of YouTube has motivated some studies related to vloggers. Not related to job seeking, Biel et al. investigated possible links between social attention received by videos and nonverbal behavior of popular vloggers [9]. They reported correlations between some nonverbal cues (e.g. speaking time,) and the average level of attention of their vlogs. In another study of YouTube vloggers, Biel et al. investigated connections between verbal content and perceived personality ratings [10]. Using audiovisual deep residual networks for annotated personality recognition, Güçlütürk et al. reported a test accuracy of 0.91 in the ChaLearn First Impressions Challenge [20]. Using this challenge dataset, Beyan et al. proposed a novel deep visual activity based features extracted only from key-dynamic images for perceived personality traits classification [8]. The authors reported their results match the state-of-the-art visual nonverbal features on average, and showed improved performance for agreeableness. Perceived hireability was investigated by Nguyen et al., who studied a sample of YouTube video resumes [39]. The authors then automatically extracted nonverbal cues and analyzed their relationship with perceived Big-Five personality and hirability variables. Muralidhar et al. extended this work to investigate verbal content and its connections to perceived hirability using manual and automatic transcripts [33].

Asynchronous video interviews have been investigated to infer various social constructs like personality, communication skills and interview ratings. Applicants' reaction to AVIs for employee selection has been a topic of research in psychology. Brenner et al. [11] integrated findings from technology acceptance research and applicant reaction to new technology research to study AVIs. Using hierarchical regression analysis of self-reported questionnaire data from $N = 106$ participants, they reported the following observations: (a) perceived usefulness and perceived ease of use were significant predictors of applicants' attitudes (b) age and gender did not predict applicants' attitudes. Langer et al. [29] investigated applicant reaction to AVIs from the perspective of media richness theories [43]. The authors used self-reported questionnaire data from $N = 113$ participants who were randomly assigned to AVI or online interview (like Skype) group. The authors reported that participants considered AVIs to be creepier ($r = .19; p < .01$), and had more privacy-related concerns ($r = .29; p < .01$) as compared to online interviews. AVIs were also perceived as less personal and have lower procedural justice than online interviews. Batrinca et al. [5] predict self-reported Big-Five personality traits in self-presentations

where participants had to introduce themselves (N=89). Participants only interacted with the interviewer for the first part of the call, while the main segment for non-verbal cue extraction was a monologue. In a similar setting, Suen et al. [54], developed an end-to-end AVI system to automatically recognize applicant personality. The authors used features automatically extracted from the interview videos and self-reported personality scores of 120 real job applicants. They reported that their interview agent recognized Big-5 personality traits of an interviewee with an accuracy between 90.9% (*Agreeableness*) and 97.4% (*Openness*). Chen et al. developed a standardized video interview protocol along with human ratings, which focused on verbal content, personality, and holistic judgment [12]. The authors using "visual words" as feature extraction method, automatically learned from video analysis outputs, and the Doc2Vec representation method resulted in a correlation of 0.42 between machine-predicted scores and human-rated scores. In another work, Chen et al. [13] investigated hirability and personality impressions using 1891 AVIs where the participant had to answer a single question. Specifically, the participants were asked to record their answer for the question: "Please tell us about a work situation in which you were not the formal leader but tried to assume a leadership role." The authors extracted various multimodal cues (i.e accoustic, visual and text), and then converted these continuous cues into discrete pseudoword documents. They then applied text classification methods in a binary classification task achieiving F-measures > 0.8 for personality traits and F-measures = 0.6 for hiring recommendation scores. Using the same dataset, Leong et al. [30] reported empirical evidence indicating biases in human annotations and automated assessment. Hemamou et al. proposed an attention neural network to infer hirability from structured video interviews (HireNet) [22]. HireNet was trained using AVIs and recruiter's annotations. In another work, the same authors studied influential nonverbal social signals in AVIs discovered using deep learning methods [21]. In contrast with their previous paper, the authors used attention mechanisms to identify the relevant parts of an answer.

So far, research in social computing has focused on developing automatic behavioral analytics systems using nonverbal and verbal cues displayed by applicants during AVIs. However, the effect of technology (like AVIs) on interviewees has not yet been explored in this community. Furthermore, psychology literature on understanding applicants' reactions to AVIs has focused on comparisons w.r.t traditional face-to-face interview methods. In this work, we address this gap and investigate applicants' reactions to AVIs using several self-reported measures from psychology research including interview performance, perception of fairness and engagement in impression management.

Figure 1: Illustration of the data collection process that includes (a) registration and consent form signing (b) pre-questionnaires (c) setup instructions and best practices (d) structure interview with 4 questions (e) post questionnaires.

3 DATASET

To answer our research questions, we collected a dataset of AVIs in French consisting of 221 participants, the majority of whom were students at a local university. This demographic is of interest given that people in this age group, might be more comfortable with the use of AVI platforms (e.g. given their frequent exposure to social media). Furthermore, this is the generation which in future will be exposed to such hiring practices and technologies. Hence, understaning their reactions to AVIs is important.

3.1 Data Collection Design

To collect AVIs and self-reported questionnaires, we designed an online platform to simulate a real job interview (Figure 1). The study was approved by the University of Lausanne ethical committee. The participants were students at a local university where the experiment was advertised and were paid the equivalent of 20 USD in local currency for the data collection which lasted about 60 minutes. The AVI platform was designed to be hosted on a web-server and was secured using 256-bit SSL certificate. The front end of the platform was written using Web technologies like HTML, CSS, and AngularJS, while the server-side was written using Node.JS and Express frameworks with a MongoDB database. A screenshot of the platform is shown in Figure 2.

3.2 Data Collection Process

To participate in the data collection, the participants had to sign in to the AVI website using their computers/laptops (but not iPads and mobile phones). They were adviced to sit in a quiet environment while completing the study.

(1) Registration & Consent: The participants went to the URL provided. Here, they first had to register using their email ids for the study. During this process, they were furnished with complete details about the study including what data would be collected and how it would be processed. They also had to agree to be video recorded by confirming an electronic consent form.

(2) Pre-questionnaires: After the participants had agreed to take part in the experiment, they were taken to the pre-questionnaires page. Here, they had to answer two questionnaires: *Big-5 Personality* and *Self Monitoring*.

(3) Setup instructions: In this section, the participants were presented with setup instructions. The instructions included essential guidelines about AVIs, some best practices like sitting in a place that was quiet and had a plain (preferably a white wall) background. The participants were then provided with a page to test their web-camera and microphone by recording a small video clip. If the participants were satisfied with the recorded video, they could proceed to the next section.

(4) Video interviews: In this section, the participants were asked the following 4 questions: Please present yourself (self-presentation); Can you describe a situation in which you took the initiative and what was the result? (initiative);Can you talk about a situation in which you managed several tasks at the same time? (organization); Give an example of a time where you had to deal with an unexpected situation that forced you to reorganize a work already planned (adaptation). The questions were displayed on the screen sequentially. For each question, the participants had 30 seconds to

Figure 2: Screenshot of the AVI platform (a)questionnaires page, (b) interview page

prepare and 4 minutes (maximum) to present the answer. After 4 minutes, a new page automatically appeared with the next question. These questions were selected from standard structured interviews by the psychology co-authors to elicit specific behavioral responses. The video interview process lasted 20 minutes [46].

(5) Post questionnaires: After the interviews, the participants had to answer eight different surveys. These included self-perceived performance, chance to perform, impression management, creepy ambiguity, privacy concerns, global fairness of the method, and perceived effectiveness. The questions were adapted from existing literature (detailed in the next section). Finally, participants were asked to provide demographic information including age, gender, current study level, recent grades, and experience with interviews (both online and face-to-face).

3.3 Pre & Post questionnaires

The details of the pre and post questionnaires are presented here. As mentioned previously, the pre-questionnaires consisted of the following measures:

(1) *Personality traits:* The participants' personality traits were measured using the Big Five Inventory (BFI-Fr)[42] on a 5-point Likert type scale ranging from 1 (strongly disagree) to 5 (strongly agree). This measure is composed of 45 items and captures the five personality traits of Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism [26]. All the items begin with "I am someone who …". Some sample questions are (a)"…is original, comes up with new ideas", (b) " …does a thorough job".

(2) *Self-monitoring (self_monitor):* This is defined as the degree to which a person observes and controls her/his expressive behavior and self-presentation in accord with social cues [19, 52]. Research in psychology indicates that the ability to self-monitor has a significant effect on relationships and professional prospects [40]. Specifically, people scoring high on this trait are characterized by sensitivity or alertness to social cues that indicate appropriate behavior. They can use such social cues to modify their self-presentation or behavior. Likewise, those scoring low on this trait are said to be relatively insensitive or less alert to social cues and tend to maintain a consistent self-presentation across different situations. In our corpus, we used the 7-item Ability to Modify Self-Presentation (AMSP) subscale to measure participants' self-monitoring ability [36]. For this, we used a 6-point Likert scale ranging from 0 (strongly disagree) to 5 (strongly agree). A sample item is "In social situations, I can alter my behavior if I feel that something else is called for".

The post questionnaires consist of the following measures:

(1) *Performance (self_rated_perf):* Participants evaluated their performance during the AVIs on a 5-point Likert type scale ranging from 1 (very bad) to 5 (very good). This self-reported performance was measured using a 3-item self-developed measure.

(2) *Chance to perform (chance_perform):* This is defined as the applicants' feeling of being given adequate opportunities to put forth their best performance [7]. The same questions posed, in face-to-face interviews and AVIs, can be viewed as providing different chances to perform. This is because, during face-to-face interviews, applicants can observe the interviewer's behavior (like head nods, smiles) and can modify their answers accordingly. This feedback is not available during AVIs by design. This lack of feedback has been shown to contribute to applicants' feeling of being offered less chance to perform during the interview [29]. In our study, participants evaluated, on a 4-item scale, the extent to which they believed that they had a chance to perform during the job interview using a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). We adapted 3 items from Langer et al. [29] and we self-developed one item. A sample item is "I could show my skills and abilities through the interview."

(3) *Impression management (IM):* This is a technique applied by applicants' in job interviews to improve their chance of being hired [46]. The IM could be honest (e.g. exaggerate their influence on positive results in past projects) or deceptive (e.g. lying about past work experience). We measured the self-reported participants' honest and deceptive IM (*honest_im* & *deceptive_im*) using 16 items adapted from Levashina et al. [31]. Each subscale is composed of 8 items. Participants indicated on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree) the extent to which they agree with each item. Sample items are "I made sure to let the interviewer know about my job credentials" and "I exaggerated my responsibilities on my previous jobs."

(4) *Creepy ambiguity (creepy_amb):* This is captured with a questionnaire about a sense of creepiness that can be elicited by unfamiliar interactions with technologies [55]. AVIs are a relatively novel method of conducting job interviews where applicants do not interact with any human interviewer but only with a system. This might be sufficient to evoke feelings of creepiness, which is defined by Langer et al. as "queasy feeling paired with uncertainty about how to behave or how to judge a situation" [27]. We collected data on the extent to which participants perceived creepy ambiguity related to the job interview method using 5 items inspired by Langer

et al. [29]. Participants indicated on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree) the extent to which they agree with each item. A sample item is "During this situation, I did not know exactly what was happening to me."

(5) *Privacy concern (privacy_concern):* This is an important aspect to investigate, given that AVIs contain sensitive personal data (like the face of the applicant and interview answers) that are recorded and transferred via the internet. These videos could also be viewed by additional interviewers, hence could result in concerns of privacy. People might feel that their privacy is being invaded through new technologies or selection procedures [53]. We collected participants' self-reports on the extent to which they were concerned about privacy issues related to the job interview method. Towards this, we used a 5-item measure adapted from Langer et al. [29]. Participants indicated their answers on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). A sample item is "Such interviews threaten applicants' privacy."

(6) *Fairness (global_fairness):* This is a measure of participants' perception of fairness regarding the selection method (AVIs) [28]. Fairness evaluations have been found to be impacted by applicants' expectations of justice. We measured global fairness using 3 items adapted from Langer et al. [29]. Participants indicated on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree) the extent to which they agree with each item. A sample item is "I think the interview procedure was fair."

(7) *Perceived effectiveness (job_pred):* We measured participants' self-reported effectiveness of AVIs to conduct the job interview using 2 items inspired by Bauer et al. [7]. Participants indicated, on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree), the extent to which they agreed with each item. A sample item is "A person receiving good evaluation on this online job interview is one that will be performing well in their job."

4 NONVERBAL BEHAVIORAL CUES

In this section, we outline the nonverbal cues extracted from the audio and visual modalities. The full list of features extracted is provided in Table 1. Several statistics, including count, mean, median, standard deviation, minimum, and maximum was computed from these cues and used as features. It must be noted here that as the dataset was collected "in the wild", there could be videos with insufficient lighting or that contained unexpected noise or background music in the audio signal.

4.1 Audio Features

Behavioral cues from speech signals can be extracted using a variety of techniques. Depending on the type of technique used for computing them, the features can be categorized into:

● **Interpretable, high-level descriptors**: these features are interpretable as they correspond to quantities used by psychologists and could be used to provide behavioral feedback. They include features like mean speaking energy. These features were computed using custom-written Python libraries.
● **Non-interpretable, low-level descriptors (LLDs)**: This set includes MFCC's, spectral centroid. The LLDs are obtained by dividing a speech signal into short overlapping frames and applying an algorithm that computes LLD at the frame level. The LLDs are then

Table 1: Detailed list of features extracted from all questions during the video interview.

Feature set	Features	Size	Ref
Interpretable (audio)	Speaking time, turn duration, speaking energy, pitch, voicing rate	35	[32, 34]
IS13 ComParE	Voice quality features (jitter & shimmer), energy, spectral, cepstral (MFCC) related LLDs, log harmonic-to-noise ratio (HNR), psycho-acoustic spectral sharpness.	6373	[50]
Visual	eye gaze duration and overall upper body motion	13	[3, 56]

converted into high-level descriptors by computing various statistics. These were extracted using the $IS-13$ ComParE configuration [50] in the OpenSmile library [17] for Python.

4.2 Visual Features

Two visual features were extracted, using OpenFace library [3], from the videos recorded during the interviews. Both these cues are interpretable (i.e. corresponds to those used by psychologists) and could be used to provide behavioral feedback. They were computed by taking a frame-wise difference in gaze angle and upper body position respectively. Facial Action Units (AUs) from OpenFace were not used as these features can be unreliable when extracted from video where participants mainly speak. Speaking affects AUs and inferred emotions, which could lead to low performance as observed in previous work in the literature [35]. Furthermore, there is increased concern on the indiscriminate use of facial expressions given their reliability [4].

● **Eye gaze**: this is an important behavioral cue in human interaction and is shown to be important in the formation of first impressions in workplaces [18, 25]. This feature is computed using the eye position and the gaze angle provided by OpenFace [3, 56]. Specifically, if the angle of gaze was less than 10°, we considering it "looking at" the camera/screen; else we consider the person to "looking away". This threshold was based on prior work [51]. Furthermore, the final value was selected by evaluating on a smaller subset of videos manually selected to have participants in varying positions in the video.
● **Upper body motion**: this is calculated as a proxy for body expressivity. This feature was reported to be correlated positively with confidence and interview performance in [49].

5 STATISTICAL ANALYSIS OF MEASURES

5.1 Descriptive Statistics

The dataset is gender balanced, with 111 males and 110 females. While the majority of the participants had no online interview experience ($N = 195$), some of the cohort had no experience even in face-to-face interviews ($N = 62$). The distribution is visualized in Figure 3.

The descriptive statistics of the self-rating for all measures is presented in Table 2. We see that the Cornbach's alpha for all is greater than 0.70 indicating high internal consistency for all measures.
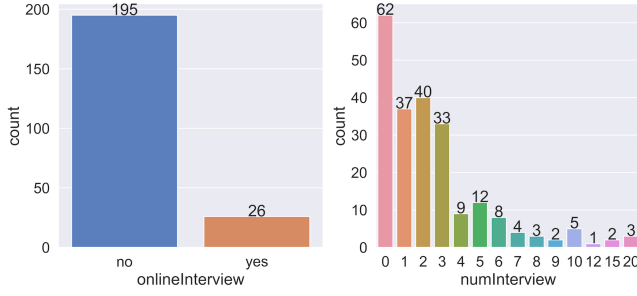
Figure 3: Participants' previous experience with (a) online interviews; (b) face-to-face interviews.

We observe that for almost all cases (except *honest_im* and *deceptive_im*), the full scale has been used. *Self-monitoring (self_monitor)* has a mean value of 4.4 indicating that participants in our corpus are sensitive to social information and might be able to use such cues to modify their own self-presentation or behavior. While the participants reported neutral impression of honest impression management (*honest_im*) (*Mean* = 2.79), they disagreed with engaging in deceptive impression management (*deceptive_im*) (*Mean* = 1.71). For all other measures, participants reported neutral impression (neither agree nor disagree) for *Chance to perform (chance_perform)*, *Creepy ambiguity (creepy_amb)*, *Privacy concern (privacy_concern)*, *Fairness (global_fairness)*, *Performance (self_rated_perf)* with mean in range [2.78, 3.15]. Mean *Perceived effectiveness (job_pred)* was low (2.15) indicating that the participants did not feel AVI performance was a good measure of how one would perform on the job.

5.2 Correlation Analysis

In this section, we present two types of correlation analysis. Note that we only discuss cases where the correlation coefficient (r) is greater than 0.2.

5.2.1 Correlation between measures. As a first step, we compute the pairwise Pearson's correlation between the self rated variables and this is shown in Table 3. We observe relatively low correlations between most of the variables, but some of them have moderate correlations. Specifically, we observe that *self_rated_perf* is moderately positively correlated to *chance_perform* ($r = .62; p < .001$) and *honest_im* ($r = .45; p < .001$). This measure is also negatively correlated to *creepy_amb* ($r = -.48; p < .001$). Furthermore, *job_pred*

Table 2: Selected descriptive statistics of questionnaire data: mean, standard deviation, min, max & Cronbach's alpha ($N = 221$).

Measures	Mean	Std Dev	Min	Max	α
self_monitor	4.40	0.77	2.43	6.00	0.85
self_rated_perf	3.15	0.73	1.33	5.00	0.71
chance_perform	2.78	0.84	1.00	5.00	0.82
honest_im	2.79	0.64	1.00	4.25	0.74
deceptive_im	1.71	0.64	1.00	4.00	0.79
creepy_amb	3.14	0.91	1.00	5.00	0.85
privacy_concern	3.14	0.83	1.00	5.00	0.77
global_fairness	2.54	0.94	1.00	5.00	0.89
job_pred	2.15	0.79	1.00	4.50	0.86

is positively correlated to *global_fairness* ($r = .29; p < .001$) and Fairness (*global_fairness*) is negatively correlated to *creepy_amb* ($r = -0.23; p < .001$). Privacy concerns *privacy_concern* is positively correlated to *creepy_amb* ($r = .25; p < .001$) and Creepy *creepy_amb* is observed to be negatively correlated to *honest_im* ($r = -.31; p < .001$) and *chance_perform* ($r = -.43; p < .001$). Honest impression management (*honest_im*) is positively correlated with chance to perform (*chance_perform*) ($r = 0.49; p < .001$).

Overall, the results indicate that participants who found the platform to be less creepy and had lower concerns of privacy rated themselves as higher performing. These participants also reported to engage in more honest impression management methods and that the platform provided them with a chance to perform well. On the other hand, participants who reported the platform as creepier, also reported more privacy concerns and rated themselves as performing less well. These participants also indicated a lower chance to perform. These results have backing in organizational psychology [11, 29].

5.2.2 Correlation between measures & nonverbal cues. In the next step, we investigated the correlations between a subset of self-reported measures (*self_rated_perf, deceptive_im* and *honest_im*) and the nonverbal behavior displayed during the video interviews. Note that in this step, we use the interpretable features only, as for the other features there would not be a plausible explanation. An interesting observation was that the behavioral cues extracted from the responses to the first question were significantly correlated to the three variables of interest. This observation has been backed up by psychology literature to some extent. Culbertson et al. reported no significant differences in behavioral cues exhibited while engaging in honest and deceptive impression management for behavioral questions (i.e. questions 2,3, and 4) [14]. Therefore, we present the results for only Question-1 (Table 4).

With this in mind, we observe that overall, all the correlations are weak, in the range ($r \in 0.16, 0.28$). We observe that participants who engaged in honest impression management had a slight tendency to speak slowly, with longer turn duration and with less pitch entropy than those who did not. Similarly, participants who engaged in deceptive impression management had a weak tendency to have lower speaking energy, with a higher pitch and speaking energy entropy, and displayed lower eye contact (gazed at screen/camera). Some of these behavioral cues are connected with deceptive impression management in the literature [15, 16, 49]. DePaulo et al. [15] in a meta-analysis combining results from 116 studies, showed that deceptive speech was less verbally and vocally involved. Specifically, people involved in deception talked less, provided fewer details, and repeated themselves more often while using a higher pitch. However, these cues had a small to moderate effect sizes (Cohen's $d = 0.25$). Furthermore, participants who rated themselves as performing higher had a weak tendency to display higher pitch, greater turn duration, spoke slowly, and had longer eye contact (gazed at screen/camera). Prior literature shows that these nonverbal cues are positively correlated to hirability in the AVI setting [13, 21, 22].

5.3 Principal Component Analysis of Measures

As a next step, we explored the possibility of obtaining a lower dimensionality representation of these variables through principal

Table 3: Pearsons correlation between all self-reported variables (N=221). (*$p < 0.001$;**$p < 0.01$;*$p < 0.05$)**

	2	3	4	5	6	7	8	9
1. self_rated_perf	0.08	0.07	−0.14*	−0.48***	−0.06	0.45***	0.62***	0.21**
2. job_pred		0.29***	0.02	−0.04	0.08	0.13*	0.16**	−0.11*
3. global_fairness			−0.15*	−0.23***	0.08	0.13*	0.27***	−0.02
4. privacy_concern				0.25***	0.13*	−0.11*	−0.18**	−0.04
5. creepy_amb					0.06	−0.31***	−0.43***	−0.16**
6. deceptive_im						0.0	0.03	−0.01
7. honest_im							0.49***	0.14*
8. chance_perform								0.06
9. self_monitor								

Table 4: Pearsons Correlation between self-reported measures of interest and nonverbal cues extracted from Q1 (self presentation) (*$p < 0.001$;**$p < 0.01$;*$p < 0.05$)**

	min pitch	mean turn duration	max turn duration	median turn duration	Q75 turns duration	std turns duration	min voicing rate	max gaze duration
self_rated_perf	−0.16*	0.21**	0.28***	0.19*	0.24**	0.17*	0.16*	0.17*

	entropy speech energy	min speech energy	Q25 speech energy	pitch entropy	mean gaze duration	max gaze duration	Q75 gaze duration
deceptive_im	0.16*	0.19*	0.19*	0.19*	−0.16*	−0.18*	−0.18*

	pitch entropy	max turn duration	min voicing rate	std voicing rate
honest_im	−0.22**	0.17*	0.16*	−0.16*

component analysis (PCA). Towards this, the self-reported variables were pre-processed to ensure that all variables had zero mean and unity variance. Then, we computed PCA using the inbuilt function in the Scikit-learn package in Python. The first two principal components (PC) account for 31.1% and 15.0% of the variance respectively. The PC is visualized in Figure 4 which displays the original variables projected onto the coordinate system.

From the PC correlation circle, we observe that *honest impression management (honest_im), Performance (self_rated_perf)* and *Chance to perform (chance_perform)* are aligned together and positively phased. While, in contrast, *Creepy ambiguity (creepy_amb), Privacy concern (privacy_concern)* are aligned together and negatively phased. These indicated the axis along which the AVIs are perceived as positive and negative by the users. Furthermore, we observe that *Fairness (global_fairness)* and *Perceived effectiveness (job_pred)* are aligned together along the PC2 indicating a global perception of this interview form.

Overall, we observe that PC1 seems to indicate an applicant's personal disposition towards AVIs (both positive and negative), while PC2 seems to indicate a global perception of AVIs including fairness for all.

6 INFERENCE ANALYSIS

We define a regression task to infer the participants self-reported values for *self_rated_perf*, *honest_im* and *deceptive_im* from automatically extracted nonverbal behavior. We evaluated two regression techniques (Ridge and Random Forest (RF)) implemented

in the "scikit-learn" package for Python [41]. The hyperparameters of the machine learning algorithms were optimized for best performance using 10-fold inner cross-validation (CV), while the performance was assessed using the 10 independent runs of leave-one-participant-out CV. We evaluated the performance of machine learning algorithms using two standard metrics: coefficient of determination (R^2) and root-mean-square error (RMSE) with baseline
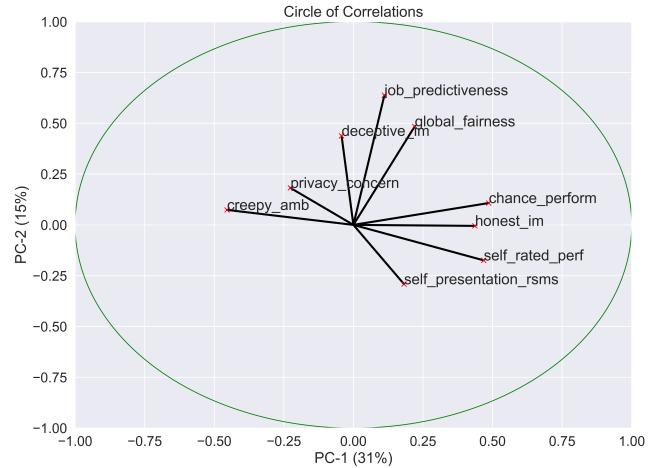

Figure 4: PCA circle of correlation displaying the original variables projected onto the coordinate system.

Table 5: Inferring self rated measures from various nonverbal feature set extracted from Q1 (self presentation).

Features	self_rated_perf		honest_im		deceptive_im		PC1	
	R2	RMSE	R2	RMSE	R2	RMSE	R2	RMSE
interpretable	0.15	0.66	0.14	0.64	0.16	0.53	0.12	2.88
non interpretable	0.12	0.52	0.16	0.65	0.17	0.53	0.13	2.76
combined	0.13	0.52	0.16	0.65	0.17	0.53	0.14	2.72

performance $R^2 = 0$. As RF outperformed ridge in all the inference experiments, we report only the results of RF here.

6.1 Inferring Self Rated Measures

In Table 5, we present the regression results. We observe that overall the performance is low for all measures of interest. Specifically, we observe that for *self_rated_perf*, the best performance is obtained using all the interpretable features ($R^2 = 0.15$) while with non-interpretable we obtain $R^2 = 0.12$. Combining the two feature sets does not improve inference performance ($R^2 = 0.13$). On the other hand, there is a slight increase in performance using non-interpretable features for honest_im and deceptive_im as compared to interpretable features ($R2 = 0.14, 0.16$ respectively). Combining the features did not improve inference results for these two measures. Inferring PC1 from the various nonverbal cues extracted too had low performance with $R^2 \in [0.12, 0.14]$. Overall, the low inference performance is in agreement with results reported in [2].

This low performance could be explained to some extent by the fact that the videos were collected in uncontolled conditions, including videos with insufficient lighting and noise in the audio signal, resulting in lower-quality extracted features. Research using similar datasets in the literature has reported lower inference performance than in lab settings. For example, Nguyen et al. [39] reported inference performance of $R^2 = 0.20$ for perceived communication skills and $R^2 = 0.15$ perceived overall impression in online video resumes compared to $R^2 = 0.36$ in face-to-face interviews [38].

7 CONCLUSIONS

In this work, we studied the reactions of applicants to the use of asynchronous video interviews. Our work adds to the understanding of this format. While most of the existing liiterature so far has focused on interviewers' (or 3rd party annotators') perspective [13, 32, 38], there has been little research on applicants perspectives in technology mediated interviews like AVIs.

Regarding RQ1, we observe two trends for reactions by participants. First, our analysis indicates that participants who found the platform to be less creepy, also had fewer concerns of privacy and rated themselves as better performing, suggesting that the platform allowed them to put their best performance and that they engaged in honest impression management. Furthermore, participants who found the platform creepy and had higher privacy concerns rated themselves as performing less well, the platform provided lower chance to perform.

Regarding RQ2, we observe that PC1 seems to indicate an applicant's personal disposition towards AVIs (both positive and negative), while PC2 seems to capture a global perception of AVIs including fairness. Regarding RQ3, we observed weak correlations between nonverbal behavior displayed and applicants perspective.

The inference performance was low, with best results for R2 in the range $[0.15, 0.17]$

Findings show that AVIs are perceived differently across young people. Results have revealed that those who do not feel at ease with this method (high creepy ambiguity score) might be penalized when participating to this type of job interviews. Indeed, these individuals rated themselves as performing less well than those who perceive positively the job interview method. From the perspective of practical use, this result shows that using AVI as an opt-in tool is a more appropriate choice for candidates, in addition to being compliant with data protection regulations like GDPR that require voluntary consent for data collection.

There are some limitations of this work. First, the AVIs were simulated interviews, hence the generalizability of the results will need to be verified in real-world situations. Second, our study involved university students applying for entry-level positions. Hence, the validity of the results should not be extrapolated to other types of applicants or positions. Third, our data collection platform followed recommendations by organisational psychologists with regard to preparation time and recording time, and was designed to be easy to use while allowing for a complex experiment. This said, the effect of certain UX choices would have to be investigated in more depth.

REFERENCES

[1] Keith Anderson, Elisabeth André, Tobias Baur, Sara Bernardini, Mathieu Chollet, Evi Chryssafidou, Ionut Damian, Cathy Ennis, Arjan Egges, Patrick Gebhard, et al. 2013. The TARDIS framework: intelligent virtual agents for social coaching in job interviews. In *Advances in computer entertainment*. Springer.

[2] Elena Margaret Lawrence Auer. 2018. Detecting Deceptive Impression Management Behaviors in Interviews Using Natural Language Processing. (2018).

[3] Tadas Baltrusaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. 2018. Openface 2.0: Facial behavior analysis toolkit. In *13th IEEE FG*. IEEE.

[4] Lisa Feldman Barrett, Ralph Adolphs, Stacy Marsella, Aleix M Martinez, and Seth D Pollak. 2019. Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological science in the public interest* 20, 1 (2019), 1–68.

[5] Ligia Maria Batrinca, Nadia Mana, Bruno Lepri, Fabio Pianesi, and Nicu Sebe. 2011. Please, tell me about yourself: automatic personality assessment using short self-presentations. In *Proc. 13th ACM ICMI*.

[6] Talya N Bauer, Carl P Maertz Jr, Michael R Dolen, and Michael A Campion. 1998. Longitudinal assessment of applicant reactions to employment testing and test outcome feedback. *Journal of applied Psychology* (1998).

[7] Talya N Bauer, Donald M Truxillo, Rudolph J Sanchez, Jane M Craig, Philip Ferrara, and Michael A Campion. 2001. Applicant reactions to selection: Development of the selection procedural justice scale (SPJS). *Personnel psychology* (2001).

[8] Cigdem Beyan, Andrea Zunino, Muhammad Shahid, and Vittorio Murino. 2019. Personality Traits Classification Using Deep Visual Activity-based Nonverbal Features of Key-Dynamic Images. *IEEE Transactions on Affective Computing* (2019).

[9] Joan-Isaac Biel and Daniel Gatica-Perez. 2011. VlogSense: Conversational behavior and social attention in YouTube. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* (2011).

[10] Joan-Isaac Biel, Vagia Tsiminaki, John Dines, and Daniel Gatica-Perez. 2013. Hi youtube!: Personality impressions and verbal content in social video. In *Proc. 15th ACM ICMI*. ACM.

[11] Falko S Brenner, Tuulia M Ortner, and Doris Fay. 2019. Asynchronous video interviewing as a new technology in personnel selection. (2019).

[12] Lei Chen, Gary Feng, Chee Wee Leong, Blair Lehman, Michelle Martin-Raugh, Harrison Kell, Chong Min Lee, and Su-Youn Yoon. 2016. Automated scoring of

interview videos using Doc2Vec multimodal feature extraction paradigm. In *Proc. 18th ACM ICMI*. ACM.

[13] Lei Chen, Ru Zhao, Chee Wee Leong, Blair Lehman, Gary Feng, and Mohammed Ehsan Hoque. 2017. Automated video interview judgment on a large-sized corpus collected online. In *7th Int. Conf. on Affective Computing and Intelligent Interaction (ACII)*. IEEE.

[14] Satoris S Culbertson, William S Weyhrauch, and Christopher J Waples. 2016. Behavioral cues as indicators of deception in structured employment interviews. *International Journal of Selection and Assessment* (2016).

[15] Bella M DePaulo, James J Lindsay, Brian E Malone, Laura Muhlenbruck, Kelly Charlton, and Harris Cooper. 2003. Cues to deception. *Psychological bulletin* (2003).

[16] Bella M DePaulo, Robert Rosenthal, Judith Rosenkrantz, and Carolyn Rieder Green. 1982. Actual and perceived cues to deception: A closer look at speech. *Basic and Applied Social Psychology* (1982).

[17] Florian Eyben, Felix Weninger, Florian Gross, and Björn Schuller. 2013. Recent developments in opensmile, the munich open-source multimedia feature extractor. In *Proc.of the 21st ACM international conference on Multimedia*.

[18] Ray J Forbes and Paul R Jackson. 1980. nonverbal behaviour and the outcome of selection interviews. *J. Occupational Psychology* (1980).

[19] Stephen J Gould. 1993. Assessing Self-Concept Discrepancy in Consumer Behavior: The Joint Effect of Private Self-Consciousness and Self-Monitoring. *Advances in Consumer Research* (1993).

[20] Yağmur Güçlütürk, Umut Güçlü, Marcel AJ van Gerven, and Rob van Lier. 2016. Deep impression: Audiovisual deep residual networks for multimodal apparent personality trait recognition. In *European Conference on Computer Vision*. Springer.

[21] Léo Hemamou, Ghazi Felhi, Jean-Claude Martin, and Chloé Clavel. 2019. Slices of Attention in Asynchronous Video Job Interviews. In *2019 8th International Conf. on Affective Computing and Intelligent Interaction (ACII)*. IEEE.

[22] Léo Hemamou, Ghazi Felhi, Vincent Vandenbussche, Jean-Claude Martin, and Chloé Clavel. 2019. HireNet: A Hierarchical Attention Model for the Automatic Analysis of Asynchronous Video Job Interviews. In *Proc.of the AAAI Conf. on Artificial Intelligence*.

[23] Annemarie Hiemstra. 2013. *Fairness in paper and video resume screening*.

[24] Mohammed Ehsan Hoque, Matthieu Courgeon, Jean-Claude Martin, Bilge Mutlu, and Rosalind W Picard. 2013. Mach: My automated conversation coach. In *Proc. ACM UBICOMP*.

[25] Andrew S Imada and Milton D Hakel. 1977. Influence of nonverbal communication and rater proximity on impressions and decisions in simulated employment interviews. *J. Applied Psychology* (1977).

[26] Oliver P John, Sanjay Srivastava, et al. 1999. The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research* (1999).

[27] M Langer and CJ König. 2017. Development of the Creepiness of Situation Scale—Study 3 convergent and divergent validity. *Retrieved from osf. io/x4umb* (2017).

[28] Markus Langer, Cornelius J König, and Victoria Hemsing. 2020. Is anybody listening? The impact of automatically evaluated job interviews on impression management and applicant reactions. *Journal of Managerial Psychology* (2020).

[29] Markus Langer, Cornelius J König, and Kevin Krause. 2017. Examining digital interviews for personnel selection: Applicant reactions and interviewer ratings. *International journal of selection and assessment* (2017).

[30] Chee Wee Leong, Katrina Roohr, Vikram Ramanarayanan, Michelle P Martin-Raugh, Harrison Kell, Rutuja Ubale, Yao Qian, Zydrune Mladineo, and Laura McCulla. 2019. To Trust, or Not to Trust? A Study of Human Bias in Automated Video Interview Assessments. *arXiv preprint arXiv:1911.13248* (2019).

[31] Julia Levashina and Michael A Campion. 2007. Measuring faking in the employment interview: development and validation of an interview faking behavior scale. *Journal of applied psychology* (2007).

[32] Skanda Muralidhar, Laurent Son Nguyen, Denise Frauendorfer, Jean-Marc Odobez, Marianne Schimd-Mast, and Daniel Gatica-Perez. 2016. Training on the Job: Behavioral Analysis of Job Interviews in Hospitality. In *Proc. 18th ACM ICMI*.

[33] Skanda Muralidhar, Laurent Son Nguyen, and Daniel Gatica-Perez. 2018. Words Worth: Verbal Content and Hirability Impressions in YouTube Video Resumes. In *Proc. of the 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*.

[34] Skanda Muralidhar, Marianne Schimd-Mast, and Daniel Gatica-Perez. 2017. How May I Help You? Behavior and Impressions in Hospitality Service Encounters. In *Proc. 19th ACM ICMI*.

[35] Skanda Muralidhar, Rémy Siegfried, Jean-Marc Odobez, and Daniel Gatica-Perez. 2018. Facing Employers and Customers: What Do Gaze and Expressions Tell About Soft Skills?. In *Proceedings of the 17th international conference on mobile and ubiquitous multimedia*. 121–126.

[36] Nils Myszkowski, Martin Storme, Franck Zenasni, and Todd Lubart. 2014. Appraising the duality of self-monitoring: Psychometric qualities of the Revised Self-Monitoring Scale and the Concern for Appropriateness Scale in French.

Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement (2014).

[37] Iftekhar Naim, M Iftekhar Tanveer, Daniel Gildea, and Mohammed Ehsan Hoque. 2015. Automated prediction and analysis of job interview performance: The role of what you say and how you say it. *Proc. IEEE FG* (2015).

[38] Laurent Son Nguyen, Denise Frauendorfer, Marianne Schmid Mast, and Daniel Gatica-Perez. 2014. Hire me: Computational inference of hirability in employment interviews based on nonverbal behavior. *IEEE Trans. on Multimedia* (2014).

[39] Laurent Son Nguyen and Daniel Gatica-Perez. 2016. Hirability in the wild: Analysis of online conversational video resumes. *IEEE Transactions on Multimedia* (2016).

[40] Olivia A O'Neill and Charles A O'Reilly III. 2011. Reducing the backlash effect: Self-monitoring and women's promotions. *Journal of Occupational and Organizational Psychology* (2011).

[41] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* (2011).

[42] Odile Plaisant, Robert Courtois, Christian Réveillère, GA Mendelsohn, and OP John. 2010. Validation par analyse factorielle du Big Five Inventory français (BFI-Fr). Analyse convergente avec le NEO-PI-R. In *Annales Médico-psychologiques, revue psychiatrique*. Elsevier.

[43] Denise Potosky. 2008. A conceptual framework for the role of the administration medium in the personnel assessment process. *Academy of Management Review* (2008).

[44] Manish Raghavan, Solon Barocas, Jon Kleinberg, and Karen Levy. 2020. Mitigating bias in algorithmic hiring: Evaluating claims and practices. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*.

[45] Sowmya Rasipuram and Dinesh Babu Jayagopi. 2016. Asynchronous video interviews vs. face-to-face interviews for communication skill measurement: a systematic study. In *Proc. of 18th ACM ICMI*.

[46] Nicolas Roulin, Adrian Bangerter, and Julia Levashina. 2014. Interviewers' perceptions of impression management in employment interviews. *Journal of Managerial Psychology* (2014).

[47] Ann Marie Ryan and Robert E Ployhart. 2000. Applicants' perceptions of selection procedures and decisions: A critical review and agenda for the future. *Journal of management* 26, 3 (2000), 565–606.

[48] Javier Sánchez-Monedero, Lina Dencik, and Lilian Edwards. 2020. What does it mean to 'solve' the problem of discrimination in hiring? social, technical and legal perspectives from the UK on automated hiring systems. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*.

[49] Leann Schneider, Deborah M Powell, and Nicolas Roulin. 2015. Cues to deception in the employment interview. *International Journal of Selection and Assessment* 23, 2 (2015), 182–190.

[50] Björn Schuller, Stefan Steidl, Anton Batliner, Alessandro Vinciarelli, Klaus Scherer, Fabien Ringeval, Mohamed Chetouani, Felix Weninger, Florian Eyben, Erik Marchi, et al. 2013. The INTERSPEECH 2013 computational paralinguistics challenge: Social signals, conflict, emotion, autism. In *INTERSPEECH*.

[51] Rémy Siegfried, Yu Yu, and Jean-Marc Odobez. 2017. Towards the use of social interaction conventions as prior for gaze model adaptation. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*. 154–162.

[52] Mark Snyder. 1979. Self-monitoring processes. In *Advances in experimental social psychology*. Elsevier.

[53] Eugene F Stone-Romero, Dianna L Stone, and David Hyatt. 2003. Personnel selection procedures and invasion of privacy. *Journal of Social Issues* (2003).

[54] Hung-Yue Suen, Kuo-En Hung, and Chien-Liang Lin. 2019. TensorFlow-based automatic personality recognition used in asynchronous video interviews. *IEEE Access* (2019).

[55] Omer Tene and Jules Polonetsky. 2014. A theory of creepy: Technology, privacy, and shifting social norms. *Yale Journal of Law and Technology* (2014).

[56] Erroll Wood, Tadas Baltrusaitis, Xucong Zhang, Yusuke Sugano, Peter Robinson, and Andreas Bulling. 2015. Rendering of eyes for eye-shape registration and gaze estimation. In *Proc. of the IEEE International Conf. on Computer Vision*.