Hire me: Computational Inference of Hirability in Employment Interviews Based on Nonverbal Behavior

4 authors, including:

Laurent Son Nguyen
Idiap Research Institute & EPFL
27 PUBLICATIONS   **896** CITATIONS

Denise Frauendorfer
University of Neuchâtel
15 PUBLICATIONS   **1,122** CITATIONS

Marianne Schmid Mast
University of Lausanne
195 PUBLICATIONS   **7,091** CITATIONS

Hire me: Computational inference of hirability in employment interviews based on nonverbal behavior

Laurent Son Nguyen, Denise Frauendorfer, Marianne Schmid Mast, and Daniel Gatica-Perez

Abstract—Understanding the basis on which recruiters form hirability impressions for a job applicant is a key issue in organizational psychology and can be addressed as a social computing problem. We approach the problem from a face-to-face, nonverbal perspective where behavioral feature extraction and inference are automated. This paper presents a computational framework for the automatic prediction of hirability. To this end, we collected an audio-visual dataset of real job interviews where candidates were applying for a marketing job. We automatically extracted audio and visual behavioral cues related to both the applicant and the interviewer. We then evaluated several regression methods for the prediction of hirability scores and showed the feasibility of conducting such a task, with ridge regression explaining 36.2% of the variance. Feature groups were analyzed, and two main groups of behavioral cues were predictive of hirability: applicant audio features, and interviewer visual cues, showing the predictive validity of cues related not only to the applicant, but also to the interviewer. As a last step, we analyzed the predictive validity of psychometric questionnaires often used in the personnel selection process, and found that these questionnaires were unable to predict hirability, suggesting that hirability impressions were formed based on the interaction during the interview rather than on questionnaire data.

Index Terms—Social computing, nonverbal behavior, hirability, employment interviews.

I. INTRODUCTION

USED in nearly every organization, the employment interview is a ubiquitous process where job applicants are evaluated by an employer for an open position. The employment interview is an interpersonal interaction between one or more interviewers and a job applicant for the purpose of assessing interviewee knowledge, skills, abilities, and behavior in order to select the most suitable person for the job at hand [45] and is one of the most popular tools to perform this task [45]. Because they require face-to-face interaction between at least two protagonists, they are inherently social [22]. As applicants and recruiters meet for the first time, employment interviews are called *zero-acquaintance* interactions [4], and all that recruiters have available as basis to forge their opinion is the applicant's verbal and nonverbal behavior during the job interview, as well as their resumes.

L. S. Nguyen and D. Gatica-Perez are affiliated jointly to Idiap Research Institute, Martigny, Switzerland, and École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland (email: lnguyen@idiap.ch; gatica@idiap.ch); D. Frauendorfer and M. Schmid Mast are affiliated to the Université de Neuchâtel, Neuchâtel, Switzerland (email: denise.frauendorfer@unine.ch; marianne.schmid@unine.ch).

In face-to-face communication the spoken words form the verbal channel, while everything else represents nonverbal communication. Nonverbal behavior can be perceived aurally (through tone of voice, intonation, and amount of spoken time, for instance) and visually (through head gestures, body posture, gaze or facial expressions) [28]. Interestingly, people quite often are able to perceive and interpret these social signals rapidly and correctly, and are often the product of an unconscious process, which makes them difficult to fake [28]. While the verbal channel remains the primary mode of communication, many social variables such as the judgment of personality, status, or competence (at the level of individuals), or the emergence of leadership or dominance (at the level of groups) are often outcomes of the multitude of micro-level nonverbal displays of behavior [28].

Nonverbal behavior in employment interviews has been studied by social psychologists for decades, mainly through the use of annotations of nonverbal cues by human observers. In the last decade, the advent of inexpensive audio and video sensors in conjunction with improved perceptual processing methods have enabled the automatic and accurate extraction of nonverbal cues, facilitating the conduct of social psychology studies. The use of automatically extracted nonverbal cues in combination with machine learning techniques has led to successful computational methods for the automatic inference of individual and group variables such as personality, emergent leadership, or dominance [19].

In this work, we present a computational framework for the automatic prediction of hirability in employment interviews. To this end, we designed and collected a dataset of 62 real job interviews and extracted audio and video behavioral features for both the applicant and the interviewer. We then used standard machine learning techniques to predict hirability scores in a regression task. To our knowledge, our work is the first one focusing on the automated prediction of employment interview outcomes from audio and visual nonverbal cues. We approach this problem from a nonverbal, face-to-face perspective, where sensing, feature extraction, and social inference are automated. The paper contains five main contributions. First, we design and collect a dataset of 62 audio- and video-recorded real job interviews, where participants were applying for a marketing job. Second, we extract audio and visual nonverbal cues related not only to the applicant, but also to the interviewer. Third, we evaluate a computational framework to infer the applicant's hirability based on the interaction during the interview. Fourth, we analyze the predictive validity of various feature groups (*e.g.* audio vs. visual cues, applicant vs. interviewer cues). Fifth, we compare the prediction performance obtained using

psychometric questionnaire data as features with the one obtained using nonverbal cues. In this work, we demonstrate the feasibility of predicting hirability to some extent, achieving to explain 36.2% of the variance.

We believe that our work is relevant for both organizational psychology and social computing. For psychologists, our study provides insights on what nonverbal cues might be used by recruiters to form the decision of hiring a person. Also, our paper shows the feasibility of using automatically extracted cues to analyze nonverbal behavior in employment interviews, as an attractive alternative to manual annotations of behavioral cues. In social computing, our research has the potential to enable the development of several applications. For instance, the findings of this study could be used for the development of a training software application for job applicants by providing them with automatic feedback on simulated job interviews rehearsed at home. Another possible application would be the development of a web-service to automatically screen job applicants, where candidates would be asked to provide, in addition to their resumes, a short video of themselves answering a series of predefined questions.

This paper is structured as follows. In Section II, we discuss the related work in organizational psychology and social computing. In Section III, we present our approach. In Section IV, we present the new data corpus collected for this study. In Section V, we discuss the methods used to automatically extract applicant and interviewer nonverbal cues. A thorough statistical analysis of the hirability scores and their relationships with behavioral cues is presented in Section VI. In Section VII, we present and evaluate the automated framework for the inference of hirability scores. In Section VIII, we analyze the predictive validity of feature groups. In Section IX, we compare the prediction accuracy obtained using nonverbal cues with the results achieved using questionnaire data. We finally conclude and discuss future work in Section X.

II. RELATED WORK

A. Related work in social psychology

In the job interview, the applicant nonverbal behavior has a remarkable impact on the hiring decision. For instance, Imada and Hakel showed that applicants who use more immediacy nonverbal behavior (*i.e.*, eye contact, smiling, body orientation toward interviewer, less personal distance) are perceived as being more hirable, more competent, more motivated, and more successful than applicants who do not [24]. Forbes and Jackson [18] showed that applicants who were employed made more direct eye contact, smiled more, and nodded more during the job interview than applicants who were rejected. Parsons and Liden [36] found that speech patterns explained a remarkable amount of variance in the hiring decision, beyond and above objective information. Also, Anderson and Shackleton [5] reported that applicants who were selected made more eye contact and produced more facial expressions during the job interview than non-accepted applicants. One explanation for the positive relation between applicant nonverbal behavior and hiring decision can be based on the immediacy hypothesis,
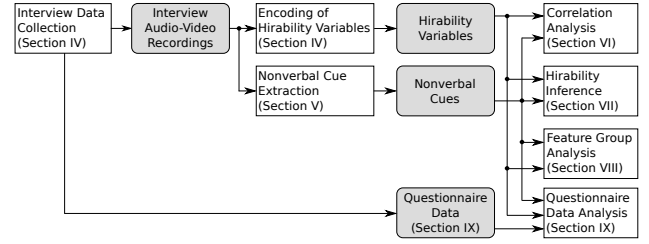

Fig. 1. Our approach

which establishes that the applicant reveals through his or her immediacy behavior (eye contact, smiling, hand gestures, etc.) a greater perceptual availability, which leads to a positive effect on the interviewer and therefore to a favorable evaluation [24].

In these studies, all coding of nonverbal behavior was done manually. Also, these works were not addressed as a prediction task in the machine learning sense (*i.e.*, no separation between training and test data was done) and the analyses were limited to correlation and in-sample ordinary least-squares linear regression.

B. Related work in social computing

Several studies have investigated computational approaches for the analysis of social constructs in face-to-face interactions from the perspective of nonverbal behavior. These automated frameworks have been used for the prediction of interest [46], dominance [25], emergent leadership [41], roles [17], end personality traits [37] [8] [10] from sensor data in small groups. Although much of the existing work investigated computational behavior analysis in small groups, some studies have also examined dyads, mainly for the prediction of outcomes in speed-dating [30] or negotiations [15] [35] interactions, but also to identify indicators of psychological disorders [42].

In the specific context of organizations, Curhan and Pentland investigated the relationship between automatically extracted audio nonverbal cues and the outcome of simulated dyadic job negotiations [15]. Related to employment interviews, Batrinca *et al.* [8] used a computational approach to predict Big-Five personality traits in self-presentations where participants had to introduce themselves in front of a computer, somewhat similar to how they would have done it in a real job interview, but without the interviewer. The authors assumed a close link between the constructs of personality and hirability; they however did not explicitly address the problem of automatic hirability prediction.

Most existing methods for the automatic inference of social constructs consist of two main steps. In the first step, behavioral features are extracted from audio (turn-taking, prosody, *e.g.* [37] [41] [25]) and video (body and head activity, visual focus of attention, *e.g.* [35]). In the second step, machine learning algorithms (including hidden Markov models [44], probabilistic graphical models [44], support vector machines [10] [8], or topic models [26]) are trained and used to predict the social constructs at hand. Our work has several points in common with these previous studies. We follow the same two-step approach, namely automatic feature extraction and

machine-learning-based inference of social constructs. However, to our knowledge this study is the first explicitly addressing the issue of inferring hirability in job interviews. Our work approaches this problem from a face-to-face, nonverbal perspective where sensing, feature extraction, and social inference are automated. Furthermore, we make use of interviewer, applicant, and relational behavioral cues extracted from both the audio and visual modalities as predictors for the regression task of inferring expert-coded hirability scores.

III. OUR APPROACH

Figure 1 displays a graphic summary of our approach. To analyze the role of interviewer and applicant nonverbal behavior in employment interviews, we first designed and collected a multimodal dataset of job interviews. Hirability scores were manually annotated by expert raters using the audiovisual recordings as basis to form their opinion. Based on the psychology literature suggesting the importance of specific nonverbal cues on the outcome of job interviews, we automatically extracted interviewer and applicant behavioral features audio and video. We then performed a correlation analysis between the extracted nonverbal cues and the hirability scores. We defined the inference task as a regression problem, where the goal was to predict the manually annotated hirability scores. We evaluated several dimensionality reduction and regression methods for the inference of hirability scores from nonverbal features. We then compared the predictive power of feature groups, *e.g.* applicant *vs.* interviewer cues, and audio *vs.* visual cues, using ridge regression as inference method. We also compared the predictive validity of questionnaire data with the performance obtained using nonverbal cues as predictors.

IV. DATA COLLECTION

For this study, we collected a corpus of 62 employment interviews. Job candidates were applying for a marketing job, where the task was to convince people on the street to participate to psychology studies. The job was paid the equivalent of 210 USD for four hours of effective work. In order to attract participants, the job was advertised using multiple communication channels among the population of three Swiss universities. Due to the large participation of students (90% Bachelor and Master students, 4.8% PhD students, 3.2% employed), the average age was 24 years ($std = 5.68$ years), and there were more female than male job applicants (45 females, 17 males).

Applicants were asked to complete a consent form in which they allowed the use of their recorded data within the framework of the project. Then, they filled in various psychology questionnaires (see Section IX for more details) before starting the job interview. The interview was designed as a structured behavioral interview, *structured* meaning that the interview strictly followed the same sequence of questions, ensuring that comparisons could be made between candidates, and *behavioral* implying that some questions were related to applicant past experiences in specific situations, eliciting a wide variety of behavioral responses. Psychology literature
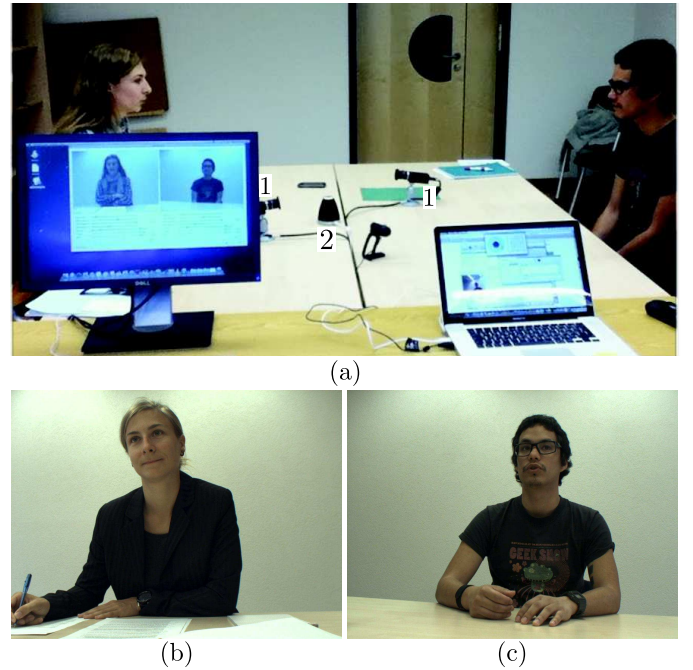
(a)

(b)

(c)

Fig. 2. (a) Snapshot of the interview room, with two HD cameras (1) and the Microcone microphone array (2). Images of the interviewer (b) and the job applicant (c) recorded by the cameras.

suggests that structured behavioral interviews are among the most valid tools for selecting applicants [23]. The interview structure is detailed in Figure 3. In total, the dataset comprises 670 minutes of recordings (average interview duration: ∼11 minutes). As an example, an interview excerpt can be seen in the **supplementary material**.

Audio and video were recorded during the employment interviews. For video, two 1280×960 cameras were used, recording both the interviewer and the job applicant synchronously at 26.6 frames per second. Camera views were quasi-frontal, filming the upper part of the body. Audio was recorded at 48kHz with a Microcone [1] microphone array placed in the middle of the table. Audio-video synchronization was done manually by adjusting the delay between the pronounced words and the lip movements. The sensor setting used in this study is illustrated in Figure 2.

Hirability is a social construct which is dependent on the type of job, the content of the interview, and how the interview is conducted [31]. For this reason, there exists no standard way of assessing hirability. Moreover, there is no single definition of hirability, but it is rather composed of several scores related to the variables that interviewers and raters assess during the interview. In this study, five hirability scores were defined, four of which were specific to the four behavioral questions of the structured interview, while the remaining one was related to the full interview sequence. More specifically, the abilities to communicate, persuade, work conscientiously, and resist stress (which were the qualities required for the job interview) were rated based on the quality of the applicant's response to the questions. Additionally, the hiring decision score was annotated on the full interview sequence (see Figure 3 for more details). Each hirability score consisted of a score

In this corpus, we used a structured behavioral design; meaning that each interview followed the same structure and that some questions were related to applicant past experiences in specific situations. The sequence of questions is listed below:

1) Short self-presentation
2) Motivation for applying for the job
3) Importance of scientific research (which is the field of the job)
4) Past experience where communication skills were required
5) Past experience where persuasion skills were required
6) Past experience of conscientious/serious work
7) Past experience where stress was correctly managed
8) Strong/weak points about self

Questions 4-7 are behavioral and were used to assess four hirability measures (Communication, Persuasion, Conscience, StressRes). Specifically, they were coded based on the quality of the applicant answers to these questions. One additional hirability measure (HirDecision) was coded on the whole interview sequence. The five hirability measures were coded by a trained psychology student on the full audiovisual recordings, where both the applicant and the interviewer were displayed.

Fig. 3. Interview structure and hirability annotations.

ranging from 1 to 5, except for hiring decision which ranged between 1 and 10. The hirability measures were annotated by a Master student in organizational psychology trained in recruiting applicants. The annotator was provided with the exact job description and job profile. She then watched the full job interviews and assigned the five hirability scores to each job applicant. For validity, a second coder (another trained organizational psychology Master student) rated 10 job interviews. Inter-rater agreement was good, with Pearson's correlation coefficient r ranging between .69 for conscience and .99 for persuasion.

Because of the privacy-sensitive content of these interviews, the dataset is not publicly available. However, the extracted features (Section V), questionnaire data (Section IX), and hirability scores can be obtained by contacting the authors.

V. BEHAVIORAL FEATURE EXTRACTION

In this study, we automatically extracted nonverbal features from the audio and visual modalities. We built multimodal and relational features by combining unimodal features. As a rationale for selecting the behavioral features to be extracted, we searched the psychology literature for nonverbal cues which were shown to play a role in job interviews. We then used available computational tools to extract the features of interest. As the interviewer's nonverbal behavior has been shown to have an impact on the interview outcome [16], we extracted behavioral cues from both the applicant and the interviewer.

A. Audio features

1) Speaking activity: Cues based on speaking activity such as applicant pauses [16], speaking time [20], and speech fluency [32] [16] were shown to have an effect on interview ratings. All speaking activity cues were based on the speaker segmentations given by the Microcone [1]. The device, in addition to recording the audio at 48 kHz, has the ability to automatically segment speaker turns, using a filter-sum beamformer followed by a post-filtering stage. The resulting speaker segmentations were stored in a file containing the

relative time (start and end) and the speaker identifier. The objective performance of the speaker segmentation was not evaluated, but upon manual inspection the number of errors was low. The following speaking-activity-based features were extracted for the interviewer and the applicant:

- *Speaking time.* Total speaking time was extracted by adding all speaking turn durations. The number was then normalized with respect to the average interview duration.
- *Speaking turns.* Speaking turns were defined as speaking segments longer than 2 seconds. Speaking turns were merged if the non-speaking gap between them was shorter than 2 seconds. The number of speaking turns, average turn duration, turn duration standard deviation, and maximum turn duration were used as behavioral features from the speaking turns.
- *Pauses.* The aforementioned non-speaking gaps shorter than 2 seconds were defined as pauses. The number of pauses were recorded, and normalized with respect to the average interview duration.
- *Short utterances.* Short utterances were defined as speaking segments of duration smaller than 2 seconds. The number of short utterances were recorded, and normalized with respect to the average interview duration.

2) Prosody: Applicant prosodic cues (*i.e.* pitch, speaking rate, and energy) were found to be significantly correlated with job interview outcomes in several psychology studies [32] [16]. From the speaker segmentations, we obtained the speech signals for the interviewer and the applicant, from which we extracted the energy, the perceived fundamental frequency, and the voiced rate (number of voiced segments per second). Methods for extracting prosodic cues are well documented (*e.g.* [7]), and we used the speech feature extraction code [2] from the Human Dynamics Group at the MIT Media Lab. For speech energy, pitch, and voicing rate, we extracted the following statistics: mean, standard deviation, minimum, maximum, entropy, median, and quartiles.

B. Visual features

Organizational psychology literature suggests that visual cues are often used by interviewers to assess the applicant's hirability in job interviews. Gaze, smiles, hand gestures, head gestures, posture, and physical attractiveness were found to have a significant effect on hirability ratings [5] [16] [20]. We decided to automatically extract a smaller number of cues including head nods, overall visual motion, and face-region optical flow. In addition to these three visual cues, we manually coded applicant gaze and smiles. In the following, we present the method used to extract these visual cues.

1) Head nods: Head nods are defined as vertical up-and-down movements of the head, rhythmically raised and lowered. We used the method proposed in [33] to automatically extract head nods, which used the Fourier transform of the optical flow in the head region, fed into a support vector machine classifier. The performance of the method was not objectively evaluated in this study. However, in a previous evaluation using a similar physical setting, the detection performance was $F_1 = 62.8\%$ at the frame level [33]. From the detected nods, we recorded

the number of nods and total nodding time. These numbers were then normalized with respect to the average interview duration.

2) Overall visual motion: This feature quantifies the amount of visual movement displayed by the applicant and interviewer during the job interview and is an indication of kinetic expressiveness. We used a modified version of motion energy images, called Weighted Motion Energy Images (WMEI) [9] which summarizes the motion throughout a video as a single grayscale image, where each pixel intensity indicates the visual activity at its position. From the WMEIs, we computed statistical features as descriptors of overall visual motion: mean, median, standard deviation, minimum, maximum, entropy, quartiles, and center of gravity.

3) Head region visual motion: This cue quantifies the amount of head motion displayed by a person, and was based on the parametric optical flow estimation method described in [34]. The overall optical flow between two consecutive frames was computed inside the face bounding box, using a parametric affine model. The estimated model was then used to compute the motion at three predefined points within the bounding box, roughly corresponding to the eyes and mouth of the person under analysis. We then took the average motion of these three points, and extracted the absolute value of the horizontal and vertical velocity components, and computed the velocity magnitude. The mean and standard deviation of these values were used as features.

4) Smiling: This cue was manually annotated by a social psychologist who counted the number of "applicant smiling" events. This number was then normalized with respect to the average interview duration. A second annotator coded smiles in a subset of the dataset ($N = 10$), and interrater agreement was high ($r = .95$).

5) Gazing: We did not use an automated method to extract this cue. An organizational psychologist coded the percentage of time for which the candidate was looking at the interviewer. A secondary annotator coded the same value on a subset of the data ($N = 10$), and interrater agreement was high ($r = .95$).

6) Physical appearance: To assess the applicants' attractiveness, three variables were coded by 10 raters based on still images: physical attractiveness, sympathy, and appreciation. Annotators were asked to answer the following questions: "How attractive do you find this person?" for physical attractiveness, "How sympathetic do you find this person?" for sympathy, and "How much do you appreciate this person in general?" for appreciation. Each rater gave a grade between 1 (low appreciation) and 5 (high appreciation), and the average over all raters was taken for the three variables.

C. Audio-visual and relational cues

Social computing studies have demonstrated the predictive validity of multimodal and relational features. For instance, features such as "looking-while-speaking" [10], cues related to the group [26] or to the dyad [15] have been successfully used for the automatic inference of social constructs. To encode the multimodal and relational characteristics of nonverbal behaviors, we combined audio and visual cues, as well as
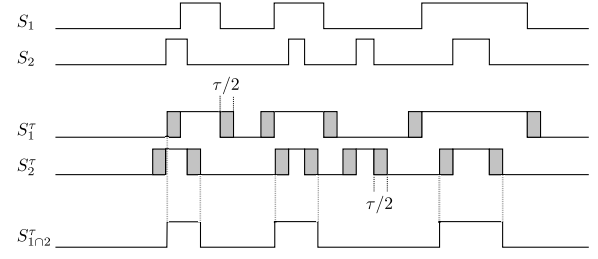

Fig. 4. Illustration of the scheme used for combining two time-series. S_1 and S_2 are the original time-series; S_1^τ and S_2^τ are the τ-dilated time-series. $S_{1 \cap 2}^\tau \triangleq S_1^\tau \cap S_2^\tau$ is the resulting time-series.

TABLE I
DESCRIPTIVE STATISTICS OF THE HIRABILITY SCORES ($N = 62$).

Score	mean	std	skew	min	max
Communication	3.016	0.983	0.489	1	5
Persusasion	3.097	1.036	-0.105	1	5
Conscience	3.097	0.953	0.379	2	5
StressRes	3.081	0.795	-0.144	1	5
HirDecision	6.161	1.803	-0.615	1	10

cues related to the applicant and to the interviewer. The rationale for combining two binary sequences is illustrated in Figure 4, and comprised two steps. First, the binary time-series were dilated using parameter τ, in order to account for the slight asynchrony between two co-occurring audio-visual or relational events. Second, the two dilated binary time-series were combined by applying a logical AND operator to each frame of the time-series. The following multimodal/relational behavioral features were extracted:

- *Audio back-channeling*: events when a person produced a short utterance while the other was speaking.
- *Visual back-channeling*: events when a person nodded while the other was speaking.
- *Audio-visual back-channeling*: events when a person nodded and produced a short utterance, using dilating parameter $\tau \in \{0, 0.5, 1, 1.5, 2\}$ seconds to account for slight asynchrony, while the other was speaking.
- *Nodding while speaking*: events when a person nodded while speaking.
- *Mutual short utterances*: co-occurring events when the two protagonists produced a short utterance, using dilating parameter $\tau \in \{0, 0.5, 1, 1.5, 2\}$ seconds to account for asynchrony.
- *Mutual nods*: co-occurring events when the two protagonists nodded, using dilating parameter $\tau \in \{0, 0.5, 1, 1.5, 2\}$ seconds to account for asynchrony.

For each of these definitions, the total time and the number of events were stored as features. The numbers were normalized with respect to the average interview duration.

VI. STATISTICAL ANALYSIS

A. Descriptive statistics of hirability scores

The descriptive statistics of the hirability scores for the full dataset are displayed in Table I. The table shows a reasonable skewness for all the hirability scores, therefore there was no need for transformation of the variables. The maximum

	1	2	3	4	5
1. Communication		$0.385^{\star\star}$	$0.261^{\star}$	$0.480^{\dagger}$	$0.590^{\dagger}$
2. Persusasion			$0.505^{\dagger}$	$0.428^{\dagger}$	$0.729^{\dagger}$
3. Conscience				$0.487^{\dagger}$	$0.630^{\dagger}$
4. StressRes					$0.699^{\dagger}$
5. HirDecision					

possible value was reached for all hirability scores, while all except conscience reached the minimal possible value. The descriptive statistics for males and females were also computed, but no noticeable differences between gender were observed, therefore these results were not reported in this study. As a last point, the normality of each hirability score was tested using the Shapiro-Wilk test; none of the variables was found to follow a normal distribution. In practice, this finding was not problematic as no assumption of normality was made in this study.

As a next step, we present the pairwise correlations (using Pearson's correlation) between the hirability variables in Table II. Note that the amount of shared variance between two variables can be obtained by taking the square of the corresponding correlation coefficient. We observe that all hirability scores were significantly correlated with each other. These correlation values suggest that the hirability scores used in this study were valid in the sense that they were measuring the same construct.

B. Analysis of nonverbal cues

As a next step, we analyzed the linear relationships between the extracted behavioral cues and the hirability measures. Pearson's pairwise correlations between the extracted behavioral cues and the hirability scores were computed. Nonverbal features which significantly correlated with hirability variables ($p < .05$) were reported in Table III.

1) Applicant behavior: We first observe that applicant cues based on speaking activity and voiced rate statistics were consistently correlated with all hirability variables. Specifically, applicants who spoke longer, faster, had longer speaking turns, and required less number of turns to answer the questions obtained better hirability ratings than candidates who did not. This finding suggests that *fluency*, *i.e.* the ability to deliver a message quickly and clearly, played a role in the formation of hirability impressions. Also, applicant who spoke longer and had longer speaking turns were perceived as more hirable, which previous psychology studies [32] [16] have already suggested a relationship between applicant fluency and employment interview outcomes, therefore our findings are supported by previous work in psychology.

To a lesser degree, applicant face region optical flow was found to be positively correlated with some hirability ratings (hiring decision and communication). Applicants who displayed more visual head motion received better hirability ratings. This observation finds some support in psychology literature [5] [20], where the amount of applicant head motion is positively correlated with interview outcomes. Similarly,

applicant statistics on WMEIs (proxy or general visual motion) were found to be positively correlated with the hirability score of persuasion. This finding also goes along the lines of previous psychology literature suggesting a link between applicant kinetic expressiveness and job interviews outcomes [5].

2) Interviewer behavior: One of the novelties of our work is the systematic study of the behavioral cues from the interviewer. Interviewer cues related to visual back-channeling, visual motion (head optical flow and some WMEI statistics), speaking activity, and prosody (voiced rate) were correlated with most hirability variables. In short, the interviewer spoke faster, had fewer speaking turns, produced more visual back-channels, and moved more in the presence of highly hirable candidates. This observation suggests that the interviewer's behavior was conditioned on the applicant: the interviewer acted differently whether she was in presence of a good or a less good job candidate. This possible instance of social mirroring between protagonists in an interaction is neither surprising nor new. An extensive body of literature (*e.g.* [28]), has demonstrated the social influence a person can have on the other protagonists of an interaction in terms of nonverbal behavior. In employment interview literature, researchers have studied the influence of the interviewer on the interview outcome by controlling his behavior (*e.g.* close *vs.* far distance, or cold *vs.* warm [29]), but to our knowledge have not specifically studied the relationship between interviewer nonverbal cues and interview outcomes. A possible hypothesis to explain our findings is that the interviewer displayed some unconscious positive behavioral responses to highly hirable applicants, by producing more visual back-channels, speaking more fluently, and showing more visual motion.

3) Mutual cues: Mutual short utterances were negatively correlated with several hirability variables (hiring decision, persuasion, and conscience). They were the only relational cues connected to hirability measures. A possible explanation comes from the fact that these mutual short utterances were in practice short back-and-forth exchanges between the applicant and the interviewer. In most cases, the applicant would ask for a clarification on the question which was just posed, such as "In my private life?", to which the interviewer would answer "As you want!". These short back-and-forth questions and answers were perceived negatively by the annotators: candidates answering questions at once without asking for clarifications received higher ratings. This finding could be related to applicant fluency: fluent candidates would answer the questions at once, without requiring further clarifications. This observation could also be related to nervousness as more nervous applicants would tend to hesitate more before answering the questions. Studying these hypotheses would require future work.

VII. INFERENCE OF HIRABILITY VARIABLES

In this section, we propose and evaluate a computational framework for the automatic inference of hirability in employment interviews. We defined the inference task as a regression problem, *i.e.* predicting the exact hirability scores, where each

TABLE III
BEHAVIORAL CUES SIGNIFICANTLY CORRELATED WITH AT LEAST ONE HIRABILITY SCORES AND CORRESPONDING PEARSON'S CORRELATION COEFFICIENT ($p < .05$, $^*p < .01$, $^\dagger p < .005$). NOT SIGNIFICANTLY CORRELATED FEATURES WERE NOT REPORTED.

Cue	HirDecision	Communication	Persuasion	Conscience	StressRes
Applicant audio cues:					
Applicant # of short utterances	$-0.477^\dagger$	-0.300	-0.307	-0.262	
Applicant speaking time	$0.528^\dagger$		0.333^*	0.307	$0.443^\dagger$
Applicant # of turns	$-0.591^\dagger$	-0.308	$-0.402^\dagger$	-0.281	$-0.451^\dagger$
Applicant avg turn duration	$0.545^\dagger$	0.257	$0.407^\dagger$	0.330	$0.439^\dagger$
Applicant turn duration std	$0.405^\dagger$		0.334^*	$0.396^\dagger$	$0.379^\dagger$
Applicant max turn duration	$0.387^\dagger$		0.301	0.355^*	$0.387^\dagger$
Applicant fundamental frequency std	-0.304				
Applicant voiced rate avg	$0.568^\dagger$		$0.365^\dagger$		0.337^*
Applicant voiced rate std	0.302		0.300		
Applicant voiced rate median	$0.451^\dagger$		0.256		0.325
Applicant voiced rate lower quartile	$0.465^\dagger$	0.286			0.354^*
Applicant voiced rate upper quartile	$0.462^\dagger$		0.321		0.283
Applicant voiced rate max	0.331^*				$0.481^\dagger$
Applicant voiced rate entropy	$0.470^\dagger$		$0.386^\dagger$		0.322
Applicant visual cues:					
Applicant vertical optical flow avg	0.330	0.328			
Applicant vertical optical flow std		0.311			
Applicant WMEI avg				0.280	
Applicant WMEI std				0.256	
Applicant WMEI lower quartile				0.256	
Applicant WMEI entropy				0.263	
Applicant WMEI vertical center of mass				0.263	
Applicant coded expressiveness				$-0.384^\dagger$	-0.302
Interviewer audio cues:					
Interviewer # of short utterances	-0.342^*			-0.343^*	-0.279
Interviewer speaking time				-0.279	-0.335^*
Interviewer # of turns	-0.349^*			$-0.367^\dagger$	$-0.425^\dagger$
Interviewer avg turn duration	0.308	0.256		0.283	
Interviewer turn duration std	0.266	$0.427^\dagger$			$0.376^\dagger$
Interviewer max turn duration		$0.380^\dagger$			0.320
Interviewer voiced rate avg	0.285				
Interviewer voiced rate median	0.281				
Interviewer visual cues:					
Interviewer optical flow magnitude avg			0.298		
Interviewer vertical optical flow avg			0.329		
Interviewer # of nods		0.271			
Interviewer nodding time	0.332^*		$0.371^\dagger$		0.279
Interviewer # of visual BC	$0.400^\dagger$	0.311	0.341^*		0.349^*
Interviewer visual BC time	$0.448^\dagger$	0.290	$0.426^\dagger$		0.343^*
Interviewer WMEI std					0.314
Interviewer WMEI max					0.333^*
Interviewer audio-visual cues:					
Interviewer # of audio video BC ($\tau = 2000ms$)				-0.315	
Interviewer audio video BC time ($\tau = 2000ms$)				-0.282	
Interviewer # of nods while speaking				$-0.385^\dagger$	
Interviewer nodding while speaking time			-0.256	$-0.415^\dagger$	
Mutual cues:					
# of mutual short utterances ($\tau = 0ms$)	-0.302				
# of mutual short utterances ($\tau = 500ms$)	-0.353^*			-0.256	
# of mutual short utterances ($\tau = 1000ms$)	$-0.387^\dagger$		-0.256		
# of mutual short utterances ($\tau = 1000ms$)	$-0.396^\dagger$				
# of mutual short utterances ($\tau = 2000ms$)	-0.302				
Mutual short utterances time ($\tau = 500ms$)	$-0.400^\dagger$			-0.302	
Mutual short utterances time ($\tau = 1000ms$)	$-0.404^\dagger$			-0.313	
Mutual short utterances time ($\tau = 1500ms$)	$-0.380^\dagger$			-0.306	
Mutual short utterances time ($\tau = 2000ms$)	$-0.366^\dagger$			-0.277	

hirability variable was considered as an independent regression task. To this end, we used a two-step approach. The first step was dimensionality reduction, and the second was regression itself, where a regression model was trained and used to predict the hirability variables.

A. Method

1) Dimensionality reduction: The goal of this step was to reduce the dimensionality of the behavioral feature vector. The feature dimensionality was not only high ($D > 140$) compared to the number of data points, it also contained by construction a large amount of redundant (*i.e.* highly inter-correlated features) and non-informative (*i.e.* cues independent of the hirability variables) data. Several standard dimensionality reduction methods were tested.

- *Low p-value features (pval).* This method assumes that the relevant information is contained in the features significantly correlated with the social variables. We selected features with $p < .05$.
- *Principal Component Analysis (PCA).* PCA is a projection onto an orthogonal space of lower dimension. It learns the linear transformation such that the variance of the projected points is maximized [27]. In this study, the

number of principal components was set such that 99.9% of the variance could be explained by the model.

- *All features (all).* In order to test the improvement of the dimensionality reduction step, we also tested the case of taking all features as predictors for the regression step.

2) Regression: In this step, the goal was to train a regression model for the prediction of the social variables. Several standard regression techniques were tested in this study.

- *Ordinary least-squares (OLS).* OLS minimizes the sum of squared errors between the observed and the predicted responses obtained using a linear model. It is the simplest regression model and is popular in psychology. The model assumes independent and identically distributed predictors, which in our study is the case only when PCA is used for dimensionality reduction.
- *Ridge regression.* Similarly to OLS, ridge regression minimizes the sum of squared errors between the observed and predicted responses of a linear model, but a regularization term is added to the cost function, which multiplies the l_2-norm of the regression coefficients. This regression penalty has the effect of shrinking the estimates towards zero, preventing the model to over-fit.
- *Random forest (RF).* Used for classification and regression, RF is based on the bootstrap aggregation of a large number of decision trees. In the regression case, standard decision trees split the feature space into hyper-cubic regions assigned to values [11]. RF aggregates the output of each separate decision tree by taking the average predicted value. RF has the advantage of being robust to over-fitting and of not making strong assumption on the input features.

We used a leave-one-interview-out cross-validation approach for training and testing the regression models. This framework used all but one interview for training, and kept the remaining one for testing. Model parameters were estimated using a 10-fold inner cross-validation approach.

B. Evaluation measures

We measured the performance of the automatic prediction models using the root-mean-square error ($RMSE$) and the coefficient of determination (R^2), as these are two widely used measures in the psychology and social computing. As the baseline regression model, we took the average hirability score as the predicted value. The $RMSE$ is defined in Equation 1, where y_{gt} are the ground truth observed variables, y_{pred} are the predicted values, and N is the number of data samples:

$$RMSE = \sqrt{\frac{\sum(y_{gt} - y_{pred})^2}{N}}. \qquad (1)$$

The coefficient of determination R^2 is based on the ratio between the mean squared errors of the predicted values obtained using a regression model and the baseline-average model. It is defined in Equation 2, where y_{gt} and $\bar{y}_{gt}$ are the observed variables and their mean; and y_{pred} are the predicted values. R^2 can be seen as the relative improvement over the baseline-average model. Note that negative value can

be obtained when the evaluated model under-performs the baseline-average model.

$$R^2 = 1 - \frac{\sum(y_{gt} - y_{pred})^2}{\sum(y_{gt} - \bar{y}_{gt})^2} \qquad (2)$$

Finally, significance levels were computed using Student's t-test on the difference between the squared residuals (*i.e.*, the difference between the predicted value and the ground truth score) of the tested regression model and the baseline-average model. The null hypothesis was defined as the mean being zero, assuming a Gaussian distribution and unknown variance. Cases where squared residuals have low average but high variance can result in low $RMSE$ and high R^2, but high p-values (*i.e.*, low significance levels).

C. Results

Table IV shows the performance of the different models for the inference of hirability variables. Performance values for OLS regression were not reported as the method consistently performed worse than the baseline-average model, due to over-fitting.

Results obtained for the hiring decision variable were significantly better than the baseline-average model for ridge regression ($p < .05$) independently of the dimensionality reduction technique, and for random forest using all nonverbal features. The best prediction result for hiring decision was obtained using ridge regression with all features ($R^2 = 0.362$), whereas using PCA for dimensionality reduction produced similar results ($R^2 = 0.360$). Hiring decision prediction results obtained with random forest were significantly more accurate than the baseline-average model when no dimensionality reduction was applied prior to the regression step ($R^2 = 0.274$), and marginally significant using low p-value features as predictors ($R^2 = 0.289$).

For the variable of stress resistance, random forest and ridge regression using low p-value features as predictors produced marginally significantly better results than the baseline-average model (respectively, $R^2 = 0.272$ and $R^2 = 0.208$). Although not statistically significant, ridge regression with other dimensionality reduction methods yielded positive results ($R^2 = 0.124$ for all features and $R^2 = 0.127$ for PCA). For persuasion, the results obtained with random forest and all features were marginally more accurate than the baseline-average model ($R^2 = 0.118$). For the remaining hirability variables (communication and conscience), no method was able to outperform the baseline average model.

D. Discussion

The results found here show the feasibility of automatically inferring the hiring decision score. Moreover, the use of nonverbal behavioral features as a basis for predicting hirability is a valid hypothesis. The variable of stress resistance was also possible to predict, even if the results were only marginally more accurate than the baseline-average model. In contrast, the variables of communication, persuasion, and conscience were more difficult to infer. A possible hypothesis to explain this finding is that raters did not form their opinion from the

cues which were extracted. Possibly, raters might rather have used more verbal content as a basis to form their opinion on these constructs than for the hiring decision. This hypothesis would have to be validated as part of future work.

To contextualize the achieved performances, we refer to the existing work in psychology. In [43], Schmidt obtained $R^2 = 0.18$ from predictors composed of nonverbal cues and a variety of "meta-behaviors" such as attentiveness, empathy, or dominance. Gifford *et al.* [20] obtained R^2 values ranging from 0.49 to 0.62 for the prediction of motivation and social skills (both perceived and self-rated), which are slightly different constructs compared to hirability. A notable exception in the literature is the work by Parsons *et al.* [36] who reported $R^2 = 0.72$; the authors of the paper themselves were surprised by this extremely high result and hypothesized that it could be an effect of the way hirability scores and nonverbal cues were annotated. In these works, R^2 results were obtained using OLS regression without separating the data into training and test sets. From this standpoint, the performance results achieved here are comparable to the ones reported in the psychology literature, with the main advantage that they were obtained using a prediction framework in the machine learning sense.

In terms of regression methods, ridge regression was the best-suited technique from the pool of methods tested for our task. The reason behind this finding may come from the fact that linear relationships between the features and the hiring decision exist, as suggested by the statistical analysis performed in Section VI; in other words, the linear assumption used in ridge regression likely held. Another interesting finding is that dimensionality reduction did not improve the prediction of ridge regression for the hiring decision score. This suggests that ridge regression was able to find the informative patterns without needing a pre-processing step, with the l_2-regularization term implicitly selecting the most informative features by assigning a low weight to redundant or uninformative features. For stress resistance, low p-value dimensionality reduction improved the accuracy, suggesting that the informative data was contained in significantly correlated features. PCA with ridge regression produced results of similar accuracy, compared with the ones obtained with no dimensionality reduction. This suggests that the transformation retained the informative data, but did not result in more informative patterns. Also, this suggests that although the predictor independence assumption was not held, ridge regression was still able to produce good predictions. On the other hand, PCA coupled with random forest showed poor prediction performance, which was not the case for the other dimensionality reduction techniques.

VIII. ANALYSIS OF FEATURE GROUPS

In this section, we analyze the predictive power of feature groups. Feature groups were defined based on the person and the modality from whom the features were extracted.

A. Method

Four groups were defined, based on the protagonist from whom the nonverbal cues were extracted: applicant, interviewer, mutual, and all. For each person-related group, the features were further separated into three sub-groups based on the modality: audio, video, and all. Based on the results obtained in Section VII showing that ridge regression with no dimensionality reduction produced in most cases the best prediction performance, the analysis of feature groups was performed using this inference method. Please note that this analysis was also done for random forest with no dimensionality reduction and yielded similar results, therefore the results were not reported. We used leave-one-interview-out cross-validation to train and test the inference method, and 10-fold inner cross-validation to select the best ridge parameter.

B. Results

The results obtained using the different feature groups as predictors are reported in Table V. Feature groups who yielded the best prediction results were audio cues extracted from both the applicant and the interviewer ($R^2 = 0.400$), interviewer visual cues ($R^2 = 0.374$), applicant audio cues ($R^2 = 0.317$), applicant audio and visual cues ($R^2 = 0.254$), and interviewer audio and visual cues ($R^2 = 0.223$).

Applicant features were predictive of the hiring decision (all applicant features: $R^2 = 0.254$, $p < .1$); results showed that the predictive applicant features stemmed from audio (applicant audio features: $R^2 = 0.317$, $p < .05$) and not from video (negative R^2). In the light of the statistical analysis conducted in Section VI, these results are not surprising as only one visual feature was found to be significantly correlated with the hiring decision. However, the result goes against related work in psychology suggesting a relationship between hirability and several visual cues (gaze, smiles, head gestures, or physical attractiveness [5]). In our case, adding the applicant visual features to the applicant audio features did not improve the prediction accuracy; rather, it decreased the performance.

Interestingly, the results obtained for the interviewer group showed good performance. In this case, interviewer visual cues showed the best accuracy ($R^2 = 0.374$, $p < .1$), whereas interviewer audio cues were not predictive. Combining audio and visual interviewer cues decreased the performance ($R^2 = 0.223$, $p < .1$) compared to the visual cues taken alone.

When grouped together, audio features extracted from the two protagonists showed the best performance across all feature groups ($R^2 = 0.400$, $p < .05$). Adding the interviewer audio cues to the applicant audio cues increased the prediction accuracy (from $R^2 = 0.317$ to $R^2 = 0.400$), even if the interviewer audio cues and the mutual audio cues were not predictive in isolation. For the group of visual cues extracted from both protagonists, the results suggest an opposite tendency: interviewer visual cues showed good accuracy ($R^2 = 0.374$), but adding the non-predictive applicant visual cues decreased the accuracy dramatically ($R^2 = 0.012$).

Also note that the significance levels are not directly linked to the R^2 and $RMSE$ values. Indeed, the applicant-audio group has lower R^2 and higher $RMSE$ ($R^2 = 0.317$, $RMSE = 1.486$) than interviewer-video ($R^2 = 0.374$, $RMSE = 1.423$), but higher significance level. These results may seem conflicting, but can be explained by the fact that the squared residuals for applicant-audio had lower variance

TABLE IV
PERFORMANCE (R^2 AND $RMSE$) FOR THE INFERENCE OF HIRABILITY SCORES USING DIFFERENT DIMENSIONALITY REDUCTION AND REGRESSION
METHODS (*$p < 0.1$, †$p < 0.05$).

Method	HirDecision		Communication		Persuasion		Conscience		StressRes	
	R^2	$RMSE$	R^2	$RMSE$	R^2	$RMSE$	R^2	$RMSE$	R^2	$RMSE$
Baseline-Avg	0.000	1.799	0.000	0.957	0.000	1.037	0.000	0.918	0.000	0.793
All-Ridge	0.362	1.437†	-0.067	0.989	0.076	0.997	-0.096	0.961	0.124	0.742
All-RF	0.274	1.533†	0.001	0.957	0.118	0.974*	0.040	0.900	0.079	0.761
Pval-Ridge	0.330	1.473†	-0.050	0.981	0.106	0.981	-0.008	0.922	0.272	0.677*
Pval-RF	0.289	1.517*	0.020	0.948	0.048	1.012	-0.036	0.935	0.208	0.706*
PCA-Ridge	0.360	1.439†	-0.067	0.989	0.069	1.001	-0.092	0.960	0.127	0.741
PCA-RF	0.079	1.726	-0.078	0.994	0.050	1.011	0.016	0.911	0.056	0.771

but higher average than for interviewer-video, resulting in a lower p-value.

C. Discussion

Applicant cues were predictive of the hiring decision score. More specifically, the relevant information stemmed from the audio modality, whereas visual features produced low prediction results. The combination of audio and visual cues decreased the performance compared to audio cues taken in isolation. One hypothesis for explaining why applicant visual cues were not predictive could be that raters used visual features which were not extracted in this study, such as applicant body posture, or fine-grain gaze patterns. The systematic examination of this hypothesis will be the subject of future work.

Interviewer cues were predictive of hiring decision, which is in our opinion an interesting finding. This suggests that by observing the interviewer only, one can to some extent infer the hirability of an applicant. This also shows that the interviewer produced behavioral responses conditioned on the quality of the applicant and that these responses were valid predictors for the hiring decision score. More specifically, the predictive validity of interviewer cues stemmed from the video modality. From Table III, the features of interest were related to interviewer visual back-channeling. Combining audio cues to visual cues however decreased the prediction performance. This finding implies that by only looking at the interviewer, it is possible to make inferences on the applicant; this approach (looking at others for inferring things about self) has been used in one previous study in a different setting [37].

When considering modalities without taking the person of interest into account, audio cues showed the best prediction performance. Interestingly, Combining applicant audio cues (high predictive validity) and interviewer audio cues (low predictive validity) actually improved the prediction performance. This finding suggests that interviewer audio cues contained some informative data, but were only useful when combined to applicant audio cues. For the visual modality, the results showed an opposite trend: combining applicant visual cues (low predictive validity) with interviewer visual cues (high predictive validity) dramatically decreased the prediction performance.

IX. ANALYSIS OF QUESTIONNAIRE DATA

The use of psychometric questionnaires for the personnel selection process is a common practice in human resources.

TABLE V
PERFORMANCE (R^2 AND $RMSE$) FOR THE PREDICTION OF THE HIRING
DECISION SCORE USING DIFFERENT FEATURE GROUPS AS PREDICTORS,
AND USING RIDGE REGRESSION WITH NO DIMENSIONALITY REDUCTION
AS INFERENCE METHOD (*$p < 0.1$, †$p < 0.05$).

Feature Group	HirDecision	
	R^2	$RMSE$
Baseline-Avg	0.000	1.799
Applicant-Audio	0.317	1.486†
Applicant-Video	-0.047	1.841
Applicant-All	0.254	1.554*
Interviewer-Audio	0.029	1.773
Interviewer-Video	0.374	1.423*
Interviewer-All	0.223	1.586*
Mutual-Audio	0.046	1.757
Mutual-Video	-0.034	1.830
Mutual-All	0.052	1.751
All-Audio	0.400	1.394†
All-Video	0.012	1.789
All	0.362	1.437†

Questionnaires are used to assess social constructs related to the task at hand. Psychology researchers have identified a number of social constructs frequently assessed during job interviews, such as intelligence, knowledge and skills, personality traits, applied social skills, interests and preferences, organizational fit, and physical attributes [23]. In this section, we analyze the predictive validity of psychometric questionnaires, in relation with hirability scores.

A. Method

During the job interview session, job applicants were asked to fill in psychometric questionnaires before starting the interview. Three types of social constructs were assessed using questionnaires:

1) *Personality*. We used the Big-Five personality model, which has received the most extensive support in psychology [21]. It represents personality at its highest level of abstraction and suggests that most individual differences in human personality can be classified into five empirically-derived bipolar factors, namely extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience (see Table VI). We used the NEO-FFI-R [14] questionnaire to assess the Big-Five personality traits. The questionnaire is standard and comprises 60 items (12 items per factor).

2) *Intelligence*. Intelligence is an important social construct in the job selection process. It has been shown to corre-

Trait	Examples of Adjectives
Extraversion	Active, Assertive, Enthusiastic
Agreeableness	Appreciative, Forgiving, Generous
Conscientiousness	Efficient, Organized, Planful, Reliable
Neuroticism	Anxious, Self-pitying, Tense, Touchy
Openness to Experience	Artistic, Curious, Imaginative

late significantly with job performance across various types of job [23]. Also, it is part of the constructs consistently assessed by interviewers [23]. We used the Wonderlic Personnel Test [3] to assess applicant general intelligence. Job applicants had to answer up to 50 questions within a time span of 12 min. The questions measured vocabulary, arithmetic reasoning, and spatial ability. The higher the number of correct answers, the higher the total score of the job applicant.

3) *Communication and persuasion*. The job for which applicants were interviewed was a marketing job, which typically requires strong social skills such as communication and persuasion. To assess these skills, a questionnaire was generated based on the Social Skills inventory [39]. Examples for items are: "In general I communicate in a clear manner" and "I often succeed in selling my point of view".

To analyze the use of questionnaires for hirability prediction, it was necessary to assume that the recruiter also had access to the questionnaire data. To this end, we performed a second round of hirability annotations, where the coder started by looking at the questionnaire outputs before watching the full interview recordings. For inter-rater agreement, a secondary coder rated a subset of the data ($N = 10$) and the agreement was good (r ranging from .72 for conscience to .93 for communication and hiring decision). For consistency, we computed the pair-wise correlations between the hirability scores based on the recordings only (used in Sections VI to VIII) and the full hirability scores; all full hirability scores were strongly correlated with their audio-video counterpart (ranging from $r = 0.839$, $p < .001$ for conscience to $r = 0.942$, $p < .001$ for hiring decision). Note that the hirability variables used in this section are the full hirability scores, *i.e.* the ones where the rater had also access to the questionnaire data. For this reason, the results can differ from the ones obtained in the previous sections.

B. Correlation analysis

Pairwise correlations between the questionnaire variables and the hirability scores are reported in Table VII. We observe that extraversion was correlated with all hirability variables except persuasion. This finding is supported by the psychology literature showing a relationship between extraversion and performance in jobs characterized by a high level of social interactions, such as in sales, marketing, or management [6]. Openness to experience was not correlated with any of the hirability variables. Psychology research has also found no strong relationship between this trait and performance or interview ratings [40]. Neuroticism was negatively correlated with

the scores of hiring decision and conscience, which also goes along the lines of related work suggesting that this characteristic negatively affects the employability of candidates [40]. No significant correlation was found for the agreeableness trait, which does not contradict the previous work in psychology as this trait was found to be related with performance only for certain occupations, such as team-work or customer service [40], which was not the type of job for which candidates were applying. Similarly, no significant correlation was found for the conscientiousness trait. This finding, however, is surprising as psychology literature showed a significant relationship between this trait and job performance across all types of occupations [6]. Also, although one of the hirability score (conscience) was specifically targeted at assessing this trait, the pair-wise correlation between the two variables was only marginal ($p < .1$). The questionnaire variable related to communication skills did not share any significant correlation with the hirability variables, even if one of them (communication) was targeted at assessing it. The same observation can be made for the questionnaire variable of persuasion. Finally, intelligence was found to share no significant correlation with the hirability scores. This observation does not match the psychology research which consistently showed a significant relationship between general intelligence and job performance across multiple types of occupations [23].

C. Prediction

To analyze the predictive validity of psychometric questionnaires with respect to hirability, we used the regression task introduced in Sections VII and VIII. We used ridge regression with no dimensionality reduction, as it consistently produced the most accurate prediction in Section VII. Furthermore, we separated questionnaire data into three groups depending on the social construct they were belonging to: personality traits, communication and persuasion skills, and intelligence. Finally, we compared the results with the ones obtained using nonverbal features, and the combination of questionnaire data with nonverbal behavior.

Results are reported in Table VIII. The prediction results achieved using questionnaire data as features were less accurate than the baseline-average model. Combining questionnaire data and nonverbal behavior did not improve the prediction accuracy compared to taking nonverbal features alone ($R^2 = 0.291$ for nonverbal behavior features *vs.* $R^2 = 0.289$ for nonverbal behavior and questionnaire data). Experiments using random forest were also conducted and the results obtained were similar to the ones produced with ridge regression, therefore were not reported for space reasons.

D. Discussion

Questionnaire data held no predictive validity for the inference of hirability variables. Even if two personality traits were found to be significantly correlated with hiring decision, they were not useful for predicting the hirability scores. In comparison, nonverbal cues produced prediction results significantly more accurate than the baseline. The use of questionnaire data put in combination with behavioral features did not improve

TABLE VII
PAIRWISE CORRELATIONS BETWEEN QUESTIONNAIRE DATA AND HIRABILITY SCORES (*$p < .05$, †$p < .01$).

	HirDecision	Communication	Persuasion	Conscience	StressRes
Extraversion	0.416†	0.274*	0.204	0.271*	0.274*
Openness	-0.038*	-0.155	0.021	-0.052	0.056
Neuroticism	-0.260*	-0.109	-0.173	-0.262*	-0.218
Agreeableness	-0.013	-0.067	-0.125	-0.066	-0.038
Conscientiousness	0.051	-0.142	0.071	0.234	0.122
Communication skills	0.094	-0.009	0.047	0.141	0.164
Persuasion skills	-0.009	-0.049	-0.046	0.014	0.100
Intelligence	0.078	0.192	0.067	-0.143	-0.046

TABLE VIII
PERFORMANCE (R^2 AND $RMSE$) FOR THE INFERENCE OF HIRABILITY SCORES USING QUESTIONNAIRE DATA AS PREDICTORS AND RIDGE REGRESSION WITH NO DIMENSIONALITY REDUCTION (*$p < .1$, †$p < .05$). RESULTS ARE THEN COMPARED TO RESULTS OBTAINED USING NONVERBAL CUES AS FEATURES.

	HirDecision		Communication		Persuasion		Conscience		StressRes	
	R^2	$RMSE$	R^2	$RMSE$	R^2	$RMSE$	R^2	$RMSE$	R^2	$RMSE$
Baseline-Avg	0.000	0.893	0.000	0.891	0.000	0.966	0.000	0.858	0.000	0.781
Personality	0.037	0.868	-0.043	0.908	-0.093	1.002	0.004	0.850	-0.064	0.792
Communication	-0.033	0.898	-0.033	0.903	-0.033	0.974	-0.068	0.880	-0.055	0.789
Intelligence	-0.038	0.901	-0.036	0.905	-0.035	0.975	-0.062	0.878	-0.033	0.781
All-Quest	-0.032	0.898	-0.026	0.900	-0.068	0.991	-0.026	0.863	-0.063	0.792
NVB	0.291	0.752†	-0.082	0.927	0.015	0.958	0.193	0.771	0.005	0.779
NVB+Quest	0.289	0.753†	-0.090	0.931	0.012	0.960	0.186	0.774	-0.006	0.783

the prediction accuracy. These findings suggest that raters used nonverbal behavior rather than questionnaire data as basis to form their opinion about the applicants' hirability. In other words, not only was nonverbal behavior more useful than questionnaire data for the prediction of the hiring decision score, but questionnaire data provided no information for inferring hirability.

Given the broad use of questionnaires in employment interviews, these results are surprising at first glance. Indeed, the results seem to contradict previous psychology research showing the validity of certain observed constructs such as intelligence or personality in the personnel selection process [23] [40]. Previous psychology studies have used personality traits as predictors for the regression of hirability or similar constructs and have reported results ranging from $R^2 = 0.16$ [12] to $R^2 = 0.43$ [13] using OLS regression. The results obtained in these works were however not obtained from a prediction task in the machine learning sense, *i.e.* there was no separation between training and test sets. We were able to reproduce results similar to [12] and [13] using the same approach (OLS regression with no cross-validation), obtaining $R^2 = 0.23$ for the hiring decision score using the Big-Five traits as independent variables, and similar results for the other hirability variables (R^2 ranging from 0.11 to 0.18). However, when separating training and test sets, we observed a drastic performance drop. This observation showcases the necessity to separate training and test sets to assess the predictive power of the independent variables of interest. It also shows that obtaining significant correlations is necessary, but not sufficient to have a reliable prediction model.

X. CONCLUSIONS AND FUTURE WORK

In this work, we proposed a computational framework for the automatic prediction of hirability in real job interviews,

using applicant and interviewer nonverbal cues extracted from the audio and visual modalities. To our knowledge, this study is the first attempt aiming at systematically analyzing nonverbal behavior in job interviews, and the first focusing on the task of hirability prediction.

We first collected an 11-hour audio-visual dataset of 62 real job interviews, where applicants were applying for a marketing job. Based on the recorded interactions, an expert rater annotated five hirability scores. Nonverbal features were then extracted for both the applicant and the interviewer, from audio and visual. As a first step, we performed a correlation analysis and found that not only applicant cues were correlated with the hirability scores, but interviewer cues, too. As a second step, we evaluated several prediction methods for a regression task. Results demonstrated the feasibility of predicting hirability scores based on automatically extracted nonverbal cues, and validated our proposed framework, with R^2 values of up to 36.2%.

We then analyzed the predictive validity of feature groups and we observed that the most predictive groups were the applicant audio cues and the interviewer visual cues. This second finding suggests that the interviewer produced behavioral responses which were conditioned on the quality of the job applicant by displaying more visual back-channels. This observation shows the potential of predicting the interview outcome by only looking at the interviewer.

As a last step, we analyzed the use of psychometric questionnaires widely used in the personnel selection process predicting hirability scores. Questionnaires were unable to predict hirability scores more accurately than the baseline model. Moreover, combining the questionnaire scores to nonverbal cues did not improve the prediction accuracy compared to nonverbal behavior only.

Several possible research directions are considered for

future work. First, more nonverbal features like postures, gestures, or fine-grain gaze patterns could be extracted and analyzed in combination with the speaking status. Secondly, we hypothesize that a significant part of the variance in the data could be explained by the verbal content. Therefore, future work could analyze the relationship between verbal content in job interviews and hirability.

Finally, while accurately predicting hirability is in our opinion relevant in and of itself, it does not tell whether the right choice was made; job performance is related to hirability, but is a different social dimension. Accurately predicting the most performing applicants using an automated method could therefore be seen as the ultimate task, but is by definition a difficult problem as many different factors which are not necessarily elicited in a job interview can affect job performance; some might even be impossible to sense. As a last point, the validity of employment interviews for selecting the most performant candidates is still an open question in organizational psychology [38], which suggests that other settings should also be considered.

ACKNOWLEDGMENTS

REFERENCES

[1] Microcone: intelligent microphone array for groups [online]. Available: http://www.dev-audio.com/products/microcone/.
[2] Speech feature extraction code [online]. Available: http://groupmedia.media.mit.edu/data.php.
[3] Wonderlic cognitive ability test [online]. Available: http://www.wonderlic.com/assessments/ability/cognitive-ability-tests/classic-cognitive-ability-test.
[4] N. Ambady, M. Hallahan, and R. Rosenthal. On judging and being judged accurately in zero-acquaintance situations. *Personality and Social Psychology*, 69(3):518–529, 1995.
[5] N. Anderson and V. Shackleton. Decision making in the graduate selection interview: A field study. *Occupational Psychology*, 63(1):63–76, 1990.
[6] M. R. Barrick and M. K. Mount. The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology*, 44(1):1–26, 1991.
[7] S. Basu. *Conversational scene analysis*. PhD thesis, 2002.
[8] L. M. Batrinca, N. Mana, B. Lepri, F. Pianesi, and N. Sebe. Please, tell me about yourself: automatic personality assessment using short self-presentations. *Proc. Int'l Conf. on Multimodal Interactions (ICMI)*, 2011.
[9] J.-I. Biel, O. Aran, and D. Gatica-Perez. You are known by how you vlog: Personality impressions and nonverbal behavior in YouTube. In *Proc. Int'l Conf. on Weblogs and Social Media (ICWSM)*, 2011.
[10] J.-I. Biel and D. Gatica-Perez. The YouTube Lens: Crowdsourced Personality Impressions and Audiovisual Analysis of Vlogs. *IEEE Transactions on Multimedia*, 15(1):41–55, 2013.
[11] L. Breiman. Random forests. *Machine learning*, pages 1–35, 2001.
[12] D. F. Caldwell and J. M. Burger. Personality Characteristics of Job Applicants and Success in Screening Interviews. *Personnel Psychology*, 51(1):119–136, 1998.
[13] M. S. Cole, H. S. Feild, W. F. Giles, and S. G. Harris. Recruiters inferences of applicant personality based on resume screening: Do paper people have a personality? *Business and Psychology*, 24(1):5–18, 2008.
[14] P. T. Costa and R. R. McCrae. *Neo PI-R Professional Manual*. Psychological Assessment Resources, 1992.
[15] J. R. Curhan and A. Pentland. Thin slices of negotiation: Predicting outcomes from conversational dynamics within the first 5 minutes. *Applied Psychology*, 92(3):802, 2007.
[16] T. DeGroot and J. Gooty. Can nonverbal cues be used to make meaningful personality attributions in employment interviews? *Business and Psychology*, 24(2):179–192, 2009.
[17] S. Favre, H. Salamin, J. Dines, and A. Vinciarelli. Role recognition in multiparty recordings using social affiliation networks and discrete distributions. In *Proc. Int'l Conf. on Multimodal Interfaces (ICMI)*, 2008.
[18] R. J. Forbes and P. R. Jackson. Non-verbal behaviour and the outcome of selection interviews. *Occupational Psychology*, 53(1):65–72, 1980.
[19] D. Gatica-Perez. Automatic nonverbal analysis of social interaction in small groups: A review. *Image and Vision Computing*, 27(12):1775–1787, 2009.
[20] R. Gifford, C. F. Ng, and M. Wilkinson. Nonverbal cues in the employment interview: Links between applicant qualities and interviewer judgments. *Applied Psychology*, 70(4):729–736, 1985.
[21] S. Gosling. A very brief measure of the Big-Five personality domains. *Research in Personality*, 37(6):504–528, 2003.
[22] J. L. Howard and G. R. Ferris. The employment interview context: Social and situational influences on interviewer decisions. *Applied Social Psychology*, 26(2):112–136, 1996.
[23] A. I. Huffcut, J. M. Conway, P. L. Roth, and N. J. Stone. Identification and meta-analytic assessment of psychological constructs measured in employment interviews. *Applied Psychology*, 86(5):897–913, 2001.
[24] A. S. Imada and M. D. Hakel. Influence of nonverbal communication and rater proximity on impressions and decisions in simulated employment interviews. *Applied Psychology*, 62(3):295–300, 1977.
[25] D. B. Jayagopi, H. Hung, C. Yeo, and D. Gatica-Perez. Modeling dominance in group conversations using nonverbal activity cues. *Audio, Speech, and Language Processing*, 17(3):501–513, 2009.
[26] D. B. Jayagopi, D. Sanchez-Cortes, K. Otsuka, J. Yamato, and D. Gatica-Perez. Linking speaking and looking behavior patterns with group composition, perception, and performance. In *Proc. Int'l Conf. on Multimodal Interaction (ICMI)*, 2012.
[27] I. T. Jolliffe. *Principal Component Analysis*. Springer-Verlag, 2nd edition, 2002.
[28] M. L. Knapp and J. A. Hall. *Nonverbal communication in human interaction*. Wadsworth, Cengage Learning, 7th edition, 2009.
[29] R. C. Liden, C. L. Martin, and C. K. Parsons. Interviewer and applicant behaviors in employment interviews. *Academy of Management Journal*, 36(2):372–386, 1993.
[30] A. Madan, R. Caneel, and A. Pentland. Voices of attraction, 2004.
[31] M. A. McDaniel, D. L. Whetzel, Frank L. Schmidt, and S. D. Maurer. The validity of employment interviews: A comprehensive review and meta-analysis. *Applied Psychology*, 79(4):599–616, 1994.
[32] T. V. McGovern. The making of a job interviewee: The effect of nonverbal behavior on an interviewer's evaluations during a selection interview. *Dissertation Abstracts International*, 37(9B):4740–4741, 1976.
[33] L. S. Nguyen, J.-M. Odobez, and D. Gatica-Perez. Using self-context for multimodal detection of head nods in face-to-face interactions. In *Proc. Int'l Conf. on Multimodal Interactions (ICMI)*, 2012.
[34] J.-M. Odobez and P Bouthemy. Robust multiresolution estimation of parametric motion models. *Visual Communication and Image Representation*, 6(4):348–365, 1995.
[35] S. Park, J. Gratch, and L. P. Morency. I already know your answer: Using nonverbal behaviors to predict immediate outcomes in a dyadic negotiation. In *Proc. Int'l Conf. on Multimodal Interaction (ICMI)*, 2012.
[36] C. K. Parsons and R. C. Liden. Interviewer perceptions of applicant qualifications: A multivariate field study of demographic characteristics and nonverbal cues. *Applied Psychology*, 69(4):557–568, 1984.
[37] F. Pianesi, N. Mana, A. Cappelletti, B. Lepri, and M. Zancanaro. Multimodal recognition of personality traits in social interactions. In *Proc. Int'l Conf. on Multimodal Interaction (ICMI)*, 2008.
[38] R. A. Posthuma, F. P. Morgeson, and M. A. Campion. Beyond employment interview validity: A comprehensive narrative review of recent research and trends over time. *Personnel Psychology*, 55(1):81, 2002.
[39] R. E. Riggio. Assessment of basic social skills. *Personality and Social Psychology*, 51(3):649–660, 1986.
[40] S. Rothmann and E. P. Coetzer. The big five personality dimensions and job performance. *Industrial Psychology*, 29(1):68–74, 2003.
[41] D. Sanchez-Cortes, O. Aran, M. Schmid Mast, and D. Gatica-Perez. A nonverbal behavior approach to identify emergent leaders in small groups. *IEEE Transactions on Multimedia*, 14(3):816–832, 2012.

[42] S. Scherer, G. Stratou, M. Mahmoud, J. Boberg, J. Gratch, A. S. Rizzo, and L.-P. Morency. Automatic behavior descriptors for psychological disorder analysis. In *Proc. Int'l Conf. on Automatic Face and Gesture Recognition (FG)*, 2013.

[43] G. F. Schmidt. *The effect of thin slicing on structured interview decisions*. PhD thesis, 2007.

[44] Y. Song, L.-P Morency, and R. Davis. Multimodal human behavior analysis: Learning correlation and interaction across modalities. In *Proc. Int'l Conf. on Multimodal Interaction (ICMI)*, 2012.

[45] W. H. Wiesner and S. F. Cronshaw. A meta-analytic investigation of the impact of interview format and degree of structure on the validity of the employment interview. *Occupational Psychology*, 61(4):275–290, 1988.

[46] B. Wrede and E. Shriberg. Spotting "hot spots" in meetings: Human judgments and prosodic cues. In *Proc. Eurospeech*, 2003.

Daniel Gatica-Perez (S01, M02) is the Head of the Social Computing Group at Idiap Research Institute and Maître dEnseignement et de Recherche at the Swiss Federal Institute of Technology in Lausanne (EPFL), Switzerland. His recent work includes computational methods to understand conversational behavior in social media, urban trends using smartphones and location-based social networks, and emerging social phenomena in face-to-face interaction. Among other professional activities, he has served as Associate Editor of the IEEE Transactions on Multimedia. He is a member of the IEEE.

Laurent Son Nguyen received his M.Sc. degree in Microengineering from the Swiss Federal Institute of Technology of Lausanne (EPFL). He is currently a Ph.D. student at EPFL and a research assistant at Idiap Research Institute. His research area of interest is the auomated analysis of human behavior in natural face-to-face interactions from sensor data, as well as to explore the effects of nonverbal behavior in employment interviews.

Denise Frauendorfer Denise Frauendorfer received her Masters degree in Psychology from the University of Zurich, in 2010. She is currently a Ph.D. student at the University of Neuchatel at the Department of Work and Organizational Psychology. Her research interests focus on interpersonal behavior and first impressions in the personnel selection process. Her recent work investigated the link between the applicants nonverbal behavior in the job interview and the applicant evaluation and job performance.

Marianne Schmid Mast received her Ph.D. in Psychology from the University of Zurich, Switzerland, in 2000. She has been a postdoctoral fellow at the Department of Psychology at Northeastern University, USA, and an assistant professor at the University of Fribourg, Switzerland. Since 2006, she is a full professor of psychology at the Depart ment of Work and Organizational Psychology at the University of Neuchâtel, Switzerland. Her research focuses on the study of interpersonal interactions, verbal and nonverbal behavior, and social perception in the realm of dominance hierarchies. Her recent work concerns social interactions and first impressions in job interview settings and the effects of power on social interactions and social perception. She currently is an Associate Editor of the Journal of Nonverbal Behavior.