A Tale of Two Interactions: Inferring Performance in Hospitality Encounters from Cross-Situation Social Sensing

SKANDA MURALIDHAR*, Idiap Research Institute, Switzerland and EPFL, Switzerland
MARIANNE SCHMID MAST, University of Lausanne, Switzerland
DANIEL GATICA-PEREZ, Idiap Research Institute, Switzerland and EPFL, Switzerland

People behave differently in different situations. With the advances in ubiquitous sensing technologies, it is now easier to capture human behavior across multiple situations automatically and unobtrusively. We investigate human behavior across two situations that are ubiquitous in hospitality (job interview and reception desk) with the objective of inferring performance on the job. Utilizing a dataset of 338 dyadic interactions, played by students from a hospitality management school, we first study the connections between automatically extracted nonverbal cues, linguistic content, and various perceived variables of soft skills and performance in these two situations. A correlation analysis reveals connection between perceived variables and nonverbal cues displayed during job interviews, and perceived performance on the job. We then propose a computational framework, with nonverbal cues and linguistic style from the two interactions as features, to infer the perceived performance and soft skills in the reception desk situation as a regression task. The best inference performance, with $R^2 = 0.40$, is achieved using a combination of nonverbal cues extracted from the reception desk setting and the human-rated interview scores. We observe that some behavioral cues (greater speaking turn duration and head nods) are positively correlated to higher ratings for all perceived variables across both situations. The best performance using verbal content is achieved by fusion of LIWC and Doc2Vec features with $R^2 = 0.25$ for perceived performance. Our work has implications for the creation of behavioral training systems with focus on specific behaviors for hospitality students.

CCS Concepts: • **Applied computing → Psychology**; • **Human-centered computing →** *Empirical studies in HCI*;

Additional Key Words and Phrases: Social computing, first impressions, hospitality, nonverbal behavior, multimodal interaction, hirability, job performance, reception desk

ACM Reference Format:
Skanda Muralidhar, Marianne Schmid Mast, and Daniel Gatica-Perez. 2018. A Tale of Two Interactions: Inferring Performance in Hospitality Encounters from Cross-Situation Social Sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 3, Article 129 (September 2018), 24 pages. https://doi.org/10.1145/3264939

1 INTRODUCTION

Nonverbal behavior is an important part of interpersonal communication and has been studied in psychology [45] and computing [31, 66]. Interpersonal communication involves not only verbal communication but also nonverbal cues like tone of voice, gestures, and facial expressions. Hence, understanding human behavior and the

*This is the corresponding author

Authors' addresses: Skanda Muralidhar, Idiap Research Institute, Martigny, Switzerland, EPFL, Lausanne, Switzerland, smuralidhar@idiap.ch; Marianne Schmid Mast, University of Lausanne, Switzerland, marianne.schmidmast@unil.ch; Daniel Gatica-Perez, Idiap Research Institute, Martigny, Switzerland, EPFL, Lausanne, Switzerland, gatica@idiap.ch.

2474-9567/2018/9-ART129 $15.00
https://doi.org/10.1145/3264939

social signals they convey is fundamental to build systems that support people communication. Most literature so far has investigated nonverbal behavior in a single setting, ranging from job interviews [17, 54, 57], behavioral feedback for improving nonverbal behavior [1, 28, 37] to workplace settings like dyadic negotiations [22] and groups at work [40, 70]. In this work, we investigate the connections between nonverbal behavior, verbal content, and first impressions in two different workplace situations: interviewing for a job and performing at the job. Specifically, we investigate links between behavior and spoken words captured through ubiquitous sensing, and impressions formed both during the job interview and the actual performance of the job itself.

The literature has shown that behavior plays a critical role in the formation of first impressions at work [45], which in turn affect the outcomes of many interactions [22, 58]. People are known to behave differently in diverse situations, as person and situation are intricately entwined. Also known as "person-situation debate" or "person-situation-behavior triad", this has been a research topic in social sciences for decades [44]. Yet, until the advent of ubiquitous computing technologies, it had been difficult to objectively quantify behavior in multiple person-situation cases, due to lack of access (both direct and unobtrusive) to interactions across situations [27, 52].

Research in organizational psychology and hospitality has studied work interactions with the aim to understand the connections between constructs like hirability and performance on one hand and behavior on the other hand. Such research has mostly relied on manual annotations until recently, making this process expensive and labor-intensive, and hence restricting the study to single situations. Advanced audiovisual sensing and analytics have been used for analysis of social constructs like hirability [17, 55, 58] and leadership [70]. Motowildo et al. studied aural and visual sources of nonverbal behavior and their correlations to performance on the job in a dataset consisting of 40 managers [53]. Similarly, DeGroot et al. evaluated the relationship between interviewees' nonverbal (visual and aural) and (a) impressions formed by the interviewers (b) interviewees job performance [24]. In both of these investigations, supervisors' ratings were considered as the measure of job performance.

In this paper, we study the connections between first impressions and automatically extracted verbal and nonverbal behavioral cues from two different situations. Specifically, we investigate connections between perceived hirability and soft skills from job interviews, behavioral cues (verbal and nonverbal) displayed during both job interviews and the reception desk, and perceived job performance and soft skills from reception desk interactions. We define *perceived performance* as the behavioral aspect of performance as perceived by others observing an interaction (like a hotel front desk, or a sale) and assessing the performance of the employee based on the interaction itself. While job performance has varied definitions in literature, our definition is derived from that proposed by Viswesvaran et al. [83], who defined job performance as "action, behavior and outcomes that employees engage in and contribute to organizational goals". We note that while specific expressions of job performance depend on the jobs and positions, some aspects can be generalized across jobs like interpersonal communication.

Towards this objective, we use a data corpus consisting of 338 videos of job interviews and reception desk interactions played by a sample of students from an international hospitality school. We address the following research questions:

RQ1: What are the connections between perceptions of candidates in job interviews and perceptions of the same person on the job?
RQ2: What is the link between automatically extracted nonverbal behavior of candidates during job interviews and the perception of performance on the job?
RQ3: What are the connections between candidates' choice of words in the two interactions and the perception of performance on the job?

To answer these questions, we use a computational framework which first extracts a rich set of nonverbal features (speaking activity, prosodic features, visual features like head nods, facial expressions using state of the art techniques) and verbal features like Linguistic Inquiry and Word Count (LIWC) and the state-of-the-art doc2vec features, and then uses machine learning methods for inference in regression tasks. Based on this framework, the contributions of this paper are:

(1) With respect to **RQ1**, we first conduct a cross-situation correlation analysis between perceived hirability and soft skills at job interviews, and perceived performance and soft skills at the reception desk. We find Pearson's correlation r in the range $[0.3, 0.49]$ implying that perceived variables in job interviews are moderate indicators of perceived performance and soft skills on the job. Second, we assess the inference of perceived performance and soft skills at the job using perceived variables in the interview setting, achieving a regression performance of $R^2 = 0.25$. The best performance ($R^2 = 0.40$) is achieved by fusing the perceived hirability and soft skills scores at the job interviews and nonverbal behavioral cues from the reception desk.
(2) With respect to **RQ2**, we first conduct a Person's correlation analysis and observe that for both interview and reception desk, specific behavioral cues (longer speaking turn and head nods) are correlated to higher ratings of all perceived variables in the corresponding situation (r in the range $[-0.43, 0.39]$). We then conduct an inference experiment to infer perceived performance and soft skills using nonverbal behavioral cues from interviews. The best performance of $R^2 = 0.30$ is obtain by fusing nonverbal cues extracted from interview and desk situations.
(3) With respect to **RQ3**, we conduct an inference experiment using linguistic content as input. We observe that the performance is lower than nonverbal behavioral cues, with best performance of $R^2 = 0.25$ using linguistic content features from reception desk setting only.

Our results have broader implications for human resources and managers in hospitality, by providing insights about potential employees' nonverbal behavior and its connections to perceived performance on the job. Our work also contributes towards building a behavioral training program across situations with a focus on hospitality students.

2 PREVIOUS WORK

2.1 Literature in Psychology

First impressions are defined as a "mental image formed of a person when met for the first time" [5]. The formation of accurate impressions from a small duration of interaction ("Thin slices") has validation in psychology [3] in various settings . Related to the workplace, Hecht and LaFrance found that thin slices of telephone operator's voice predicted job performance as measured by the shortness of their calls [34]. Hall et al. showed that patient satisfaction with doctors can be predicted using this slices of the doctor's voice rated on anger, dominance, sympathy [33]. Regarding assessment of performance, it has been shown that end of semester ratings of teachers could be inferred from 10-second clips [4]. Ambady et al. reported the predictive validity of thin slice judgments, using 30-sec audio clips, on the performance of 12 sales managers [2].

Other literature has shown that nonverbal behavior plays an important role in determining the outcome of job interviews. Early works reported that applicants who displayed eye contact, smiling, etc. were perceived to be more hirable than applicants who did not [41]. Other works reported similar correlations between visual cues and impressions of hirability [6, 26]. Fluency of speech and other nonverbal speech cues have been shown to explain hirability scores [25, 51, 62].

Verbal content and its contribution to outcomes of job interviews have also been investigated [35, 36]. In a study consisting of 338 on-campus job interviews [36], Hollandsworth et al. reported that appropriateness of content contributed to favorable outcome of job interviews. Other variables contributing to this were fluency of speech and composure. Similar results were reported in a study consisting of 80 simulated interviews of undergraduate students [68], showing that relevant verbal content along with consistent nonverbal behavior had the highest impact on positive interview outcomes. Kuncel et al., in a meta-analysis consisting of MAT (Millers Analogies Test) and Graduate Record Examination (GRE), reported that measures of MAT scores were valid predictors of academic and vocational performance and evaluations of career potential and creativity [47], suggesting that intelligence at school and at work are not wholly different.

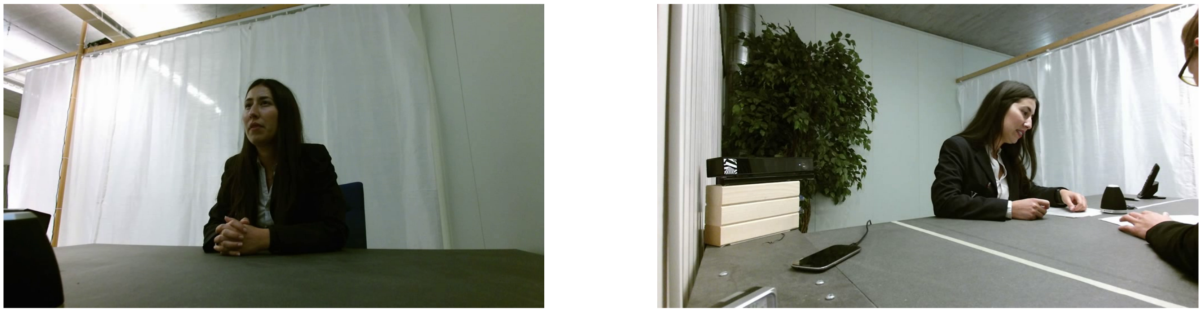
In hospitality and marketing, face-to-face interaction between service employees and customers is a critical element of customers' perceptions of the service organization. Research in this domain has demonstrated the interplay between behavior and perceived quality of service. Gabbott and Hogg recorded video of an actress playing a reception desk assistant [29] tasked with helping a customer check-in at a hotel. The study, using a 377 questionnaires answered by students, showed that nonverbal communication impacts perceived quality of service. The effect of nonverbal communication styles on customer-oriented service performance was investigated by Kang and Hyun in a study consisting of 527 luxury restaurant patrons [43], who completed a questionnaire about their recent visit to a luxury restaurant in the United States. The study reported that customer satisfaction was positively correlated to communication styles that displayed greater immediacy behavior like smiling, nodding etc. The role of nonverbal behavior in customer satisfaction was investigated by Jung and Yoon in a study of 333 customers at a family restaurant in South Korea [42]. The authors investigated the correlation between customer satisfaction and various nonverbal cues (visual and paralingual), and found a positive correlation between visual nonverbal cues (gestures, head nods) and customer satisfaction ($r = 0.42; p < 0.01$), and between paralingual nonverbal cues and customer satisfaction ($r = 0.33; p < 0.01$).

Due to the difficulty in obtaining direct behavioral measurement across multiple situations, it has been traditionally challenging to quantify behavior in "person-situation" interactions [27]. There are few works in psychology that have investigated impressions and behavior across multiple situations, especially those investigating behavior displayed during job interviews with performance on the job. Motowildo et al. investigated the connection between aural and visual cues displayed in a structured interview of 40 managers from a utility company with the performance ratings of these managers by their supervisors [53]. The recorded interviews were rated by 194 undergraduate students on the same scale. The study reported a correlation of $r = 0.36$ between the student ratings and supervisors ratings of performance. This work was extended by DeGroot et al. who investigated various nonverbal cues and their correlation to performance ratings [24]. This work used a dataset of 110 managers from a news-publishing company, and reported that vocal cues (pitch, pitch variability, speech rate, and pauses) correlated to performance ratings ($r = 0.20, p < 0.05$). The study also reported that visual cues (physical attractiveness, smiling, gaze, hand movement, and body orientation) had low correlation to performance ratings ($r = 0.14, p < 0.05$).

All of the above research so far relied on manual behavior coding. This process is expensive and laborious, and hence it is difficult to investigate many features or multiple situations, making such studies rare in the literature [27]. This situation has advanced through ubiquitous technologies.

2.2 Literature in Computing

In job interview situations, Batrinca et al., using a computational approach to infer Big-5 personality traits from short self-presentations video clips ($30 - 120$ seconds), where participants introduced themselves in front of a computer [7]. This setting is similar to job interviews, but without the presence of an interviewer. Nguyen et al., using a dataset of 62 real job interviews, automatically extracted nonverbal cues (speaking turns, prosody, head nods, visual activity) from applicant and interviewer to infer five hirability variables with up to $R^2 = 0.34$ [58]. Naim et al., analyzed 138 simulated job interviews by extracting various nonverbal cues (facial expressions, verbal content, and prosody) to infer several variables including hiring recommendation [57]. A standardized video interview protocol was proposed by Chen et al. [17], which focused on verbal content and "visual words" to predict scores of personality, and a holistic judgment. They achieved a correlation of 0.42 between machine-predicted scores and human-rated scores on a data corpus of 36 videos. In a previous work using a dataset of 169 interviews,



(a) (b)

Fig. 1. Snapshot of the situations studied in this work: (a) job interview, (b) reception desk

we found that various nonverbal cues like speaking activity, prosodic and visual activity were correlated to perceived soft skills and overall impression [55].

In the context of job interviews, the impact of linguistic style has also been investigated [17, 54, 57]. Naim et al., studied a dataset of 138 undergraduate students and reported that certain linguistic cues were linked to positive hirability scores [57]. The authors, using Linguistic Inquiry and Word Count (LIWC) to extract linguistic style, reported that the use of "We" instead of "I", more unique words, and fewer fillers words connected to positive impressions. Chen at al., reported improved prediction of expert scores using Doc2Vec algorithm and LIWC in a dataset consisting of 36 simulated job interviews [17]. Doc2Vec is a numerical representation of words popular in natural language processing [48]. The study reported that a combination of Doc2Vec and LIWC explains up to 20% of variance in the expert ratings of interview performance.

Job performance in organizational psychology has various definitions. In computing, a few constructs have been investigated in settings other than job interviews. Connections between successful negotiation and audio cues (voice activity levels, prosodic emphasis, and vocal mirroring) were investigated by Curhan et al. [22]. The negotiated compensation package was the measure of performance and the audio features explained up to 30% of the variance. In another work Lepri et al. [49], individual performance in a group task was inferred using nonverbal behavior. Raducanu et al. investigated the use of nonverbal behavior to predict the participant to be fired in the reality TV show "The Apprentice" [67]. The authors studied speaking turn features and predicted the candidate to be fired with an accuracy of 92%. In the context of hospitality encounters, we used automatically extracted nonverbal cues to infer impressions of performance and skills [56], obtaining a performance of $R^2 = 0.30$ using only nonverbal cues, and $R^2 = 0.37$ by fusing nonverbal cues and personality impressions.

The existing literature has demonstrated the feasibility of inferring some social constructs using thin slices of nonverbal behavior extracted for a single situation. In contrast to all the discussed works, we investigate the inference of perceived performance and soft skills using automatically extracted nonverbal cues across situations. To the best of our knowledge, there has been no ubicomp studies of nonverbal behavior and verbal content in multiple situations in the context of workplaces. We believe, therefore, that this work could have wider implications for behavioral training programs, human resources, and hospitality.

3 DATASETS

We used a data corpus consisting of 169 interactions each in two situations; job interview and reception desk, previously collected by our team [55]. This data was collected as a part of a behavioral training program designed for an international hospitality school. This section briefly outlines the process of data collection and annotation of perceived variables.

3.1 Data Collection

The data collection process, described in [55], consisted of two lab sessions. Each session consisted of two roles: one job interview and one front desk interaction, each played between the protagonist and a team of trained research assistants playing the roles of interviewers and clients.

The corpus consists of 100 students from the hospitality school who took part voluntarily. 69 students participated in the second session, while 31 did not return. The mean age of participants was 20.6 years, with 57 females and 43 males. The interactions were in either English or French (based on the choice of each participant) due to the international nature of the school, resulting in 260 (resp. 78) interactions in French (resp. English). Overall, the job interview dataset is 1690 minutes long (mean duration: 10 mins), while the reception desk dataset is 1350 minutes long (mean duration: 8 mins). In our investigation, we use the entire 338 videos (169 videos from each setting) and analyze at video-level. The two lab sessions were recorded $4-6$ weeks apart. So we treat them as independent videos in line with ubicomp literature [37].

Both lab sessions were captured with multiple modalities. The video data of the interactions was recorded using two Kinect v2 devices (one for each interaction partner), and was recorded at 30 fps in RGB and depth (1920×1080 and 512×424 for RGB and depth, respectively.) Audio data was captured at 48kHz with a microphone array device that segmented speaker turns from localized sources. Audio and video streams are synchronized. A snapshot of the corpus is presented in Figure 1a,1b.

3.2 Scenario for Each Situation

3.2.1 Job Interview. The job interview situation consists of the participant applying for an internship at a high-end hotel (Figure 2a). A structured interview process, with each interview following the same sequence of questions, was employed. This process has been shown in psychology to be among the best tools to compare interviewees and select applicants [39]. The interviews were conducted by a pool of seven trained research assistants who were university students in organizational psychology and business.

3.2.2 Reception Desk. The reception desk situation consists of a role play between a receptionist (the participant) and a client (one of the research assistants). The participants were informed of the situation before starting (Figure 2b) but are unaware of the client's exact reaction. The aim of the situation was to assess participants' ability to handle an unfriendly client in the best possible manner. The scenario was slightly modified in the second lab session to reduce its predictability. In this session, the participants interact with a new client (i.e., a different research assistant) who changes her/his attitude even before receiving the bill by complaining about a bad restaurant recommendation by a previous receptionist.

3.3 Annotations

The data was augmented with a number of manually labeled variables as described in [55]. The job interview videos were annotated by a group of five independent annotators, while the reception desk videos were annotated by a different group of three independent raters. Both groups of annotators were students and paid 20 CHF per hour for their work. The annotators in both groups rated the videos on various perceived variables on a seven-point Likert scale after watching the first two minutes of the videos (self-presentation in the job interview and complaint segment in the reception desk).

The perceived variables annotated for both situations along with their descriptive statistics are listed in Table 1. We use Intraclass Correlation Coefficient (ICC) [74], to measure the agreement between raters. Specifically, $ICC(2, k)$ is used as the measure of the inter-rater agreement because a sample of annotators was used, and each

Interview Questions

1. Short self-presentation.
2. Motivation for working in the service industry.
3. Past experience requiring attention to details.
4. Past experience where stress was correctly managed.
5. Past experience where adaptability was required.
6. Past experience of calm and tact under stress.
7. Strong/weak points about self.

(a)

The participant is a reception desk intern at a high-end hotel. The manager is unavailable & the participant has to handle all customer interactions. A client comes to check out. The initial interaction is friendly. The customer's attitude becomes unfriendly after the bill is recieved and starts complaining about issues faced during the stay. The participant needs to handle the situation and come to a resolution that is acceptable to the client.

(b)

Fig. 2. Details of each situation during data collection (a) Questions asked during job interview where the participant is role playing an applicant for an intership in a high-end hotel (b) Reception desk setting in which participants have to handle an unhappy client.

annotator judged all videos. From Table 1, we observe that the agreement among raters for all perceived variables was moderate to high with $ICC(2,k)$ in the range $[0.52, 0.77]$ for interview videos, while for the reception desk the $ICC(2,k)$ is in the range $[0.60, 0.77]$. ICC values greater than 0.5 are generally considered to be acceptable inter-rater agreement. For both situations, the distribution of all the perceived variables are centered on the positive side of the Likert scales (Mean ≥ 4) implying that both groups of annotators generally perceived the participants positively.

3.4 Speech Transcripts

To investigate the impact of linguistic content employed by participants in each situation, we used manually transcribed text from the audio tracks. We choose manual transcription instead of using an automatic speech recognition (ASR) system to set a gold standard against which future ASR works could be compared with. The transcription was done by a pool of five master's students in organizational psychology, who were native French speakers and fluent in English, watched all the videos, and transcribed the interaction in the original language. The transcribed documents contained verbal content of both the research assistants' and the participants' speech.

Table 1. List of perceived variables manually annotated for both situations, along with their $ICC(2,k)$ and means.

	Job Interview		Reception Desk	
Variable	$ICC(2,k)$	**Mean**	$ICC(2,k)$	**Mean**
Professional Skills				
Competent (compe)	0.56	6.01	0.69	4.24
Motivated (motiv)	0.52	5.89	0.63	4.80
Social Skills				
Positive (posit)	0.60	5.70	0.60	4.34
Sociable (socia)	0.57	5.67	0.64	4.46
Communication Skills				
Clear (clear)	0.67	5.89	0.66	4.56
Persuasive (persu)	0.69	5.57	0.72	4.01
Overall				
Performance (peImp)	–	–	0.77	4.11
Hirability (hire)	0.69	5.54	–	–

In our analysis, we use only the participants' data for two reasons: our focus is on participants behavior, and the research assistants' questions did not vary during the job interview situation.

The average number of words for an interview (applicant answers only) was 813, with a minimum of 358 and a maximum of 2587 words. For the desk situation, the mean number of words was 354.1, with a minimum of 140 words and a maximum of 1027 words. This difference in the mean number of words is due to the scenario, as the job interview setting needs the participant to speak more, while in the reception desk interaction, the client is unhappy and speaks more.

4 NONVERBAL AND VERBAL FEATURE EXTRACTION

A number of features were extracted to characterize the nonverbal behavior and verbal content of participants. The choice of nonverbal cues and verbal features was guided by existing literature in psychology [2, 19, 24, 41] and computing [17, 49, 55, 58]. The following cues are extracted:

1. **Acoustic Features** are divided into two types: speaking activity and prosody. Studies in interpersonal communication indicate that vocal characteristics such as pauses, pitch, and loudness are used by listeners to perceive the speakers' intent [72, 75]. The feature vector of this modality is of length 98.
   a **Speaking Activity Features** have been shown in psychology and hospitality literature to be correlated to impression formation in various workplace interactions [12, 24], and have been validated in the computing literature [49, 58]. We extract various features based on the speaker segmentation provided by a commercial microphone array. These include speaking time (total time that an individual speaks), speaking turns (active segments greater than two seconds), pauses (gaps in speech shorter than two seconds), short utterances (speaking segments shorter than two seconds), and silence (gaps in speech greater than two seconds).
   b **Prosody Features** were extracted from freely available MATLAB code [16, 65]. These features include pitch (voice fundamental frequency), speaking rate (speed at which words are spoken), spectral entropy (measure of irregularity or complexity), energy (voice loudness), voicing rate (number of voiced segments per second), and time derivative of energy (voice loudness modulation). The following statistics were extracted and used as features: mean, standard deviation, minimum, maximum, entropy, median, and quartiles.
2. **Visual Features** are further divided on the basis of overall body motion, head nods, and facial expressions. These cues play an important role in nonverbal communication and have been shown to influence interviewer's and clients' assessments [38, 51, 80]. Gesturing contributes to the effectiveness of a message being delivered [13], while head nodding enhances perceptions of empathy, courtesy and trust [80]. Facial expressions are associated with various social behavior including dominance [60, 79], warmth [8] and emotional distance [15]. A number of statistics, including count, mean, median, standard deviation, minimum, maximum, entropy, quartiles, and center of gravity, were computed as features. The length of the visual feature vector is 64.
   c **Overall Visual Motion** captures the total amount of visual movement displayed during the entire interaction. This feature is computed by a modified version of motion energy images, called Weighted Motion Energy Images (WMEI) [9].
   a **Head Nods & Channeling:** Head nods were extracted using a 3D face centered method [18]. In this method, a 3D head tracker calculates the angular velocities using relative rotation at each instant with respect to the head pose at an earlier instance. This method provided a per frame output, with nodding indicated with 1 and no nods indicated by -1. We define visual back-channeling (visual BC) as an event when a person nodded while the other was speaking. This cue was obtained by synchronizing speaking activity with head nod activity. Another cue we extracted is nodding while speaking.

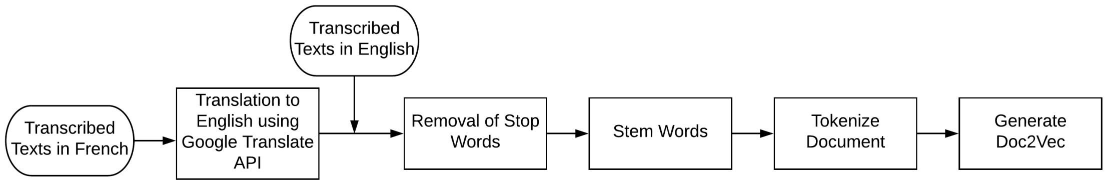


Fig. 3. Flow chart showing the steps followed for extracting Doc2Vec features from transcribed text of both the situations.

b **Facial Expression Features** were extracted using the Emotion API of Microsoft Azure cognitive services [77, 78]. Various cloud-based services [73] are available and has been previously used in the literature to study diverse social, political, and information interaction issues including cyber-bullying [76] and public health images [30]. In this work, we use the Microsoft Azure Emotion API [20] to extract emotions from facial expressions. As a first step, video frames were extracted from video clips at 5 frames per second (fps). Then, these images were input sequentially to the API. The output was confidence values across 8 facial expressions of emotion (happiness, sadness, surprise, anger, fear, contempt, disgust and neutral) normalized to 1. If a face was not found, the API returned 0 for all values and were filtered before processing. Various statistics were computed from this 8-dimensional vector and used as features.

3. **Linguistic Features:** Peoples' choice of words while speaking and writing reveal aspects of a person's identity [19] while also providing cues to their thought processes, emotional states, intentions, and motivations [64, 81]. Although the impact of linguistic content on perceived hirability and soft skills have investigated in the literature [17, 54, 57], less is known about the role of linguistic content on perceived performance.

a **Lexical Features** were extracted using Linguistic Inquiry and Word Count (LIWC) [64], a software module widely used in social psychology [19] and social computing [10, 71]. LIWC looks up each word in the interview transcript to the dictionary, and then maps them to one of 71 categories (e.g proper pronouns, adjectives, verbs etc) and increments the appropriate word category. It must be noted that LIWC can assign words to more than one category at a time. After processing a document, LIWC divides the count of categories by the total number of words in the document. LIWC is designed to process raw text, thus no pre-processed of transcripts was required. The total length of this features is 70.

c **Doc2Vec** is a Python implementation [69] of the paragraph vector method [48]. Doc2Vec generates a fixed length vector for numerical representation of text data of varying size, such as sentences, paragraphs, or documents. The steps taken to extract Doc2Vec features is summarized in Figure 3. Our data corpus consists of two languages (French and English) so as a first step we translated all the French text into English. This translation was done using the Google Translate API, a translation system in which language models are trained on billions of words of equivalent text in different languages. This API was found to be the most accurate in three of four tests consisting of 20 machine translation systems [59]. This is a necessary step as otherwise the word embedding trained would be in two separate spaces making any comparisons difficult. We then remove stop words from the text data [82]. In the next step, the text is converted into lower case, stemmed and tokenized using the NLTK package [11] in Python. We then generate document vectors by training a model for word embedding using the Gensim package [69]. The model was generated by selecting a constant learning rate for 10 epochs with 100 iterations and a vector of length 100. The choice of these parameters was guided by the small size of our data.

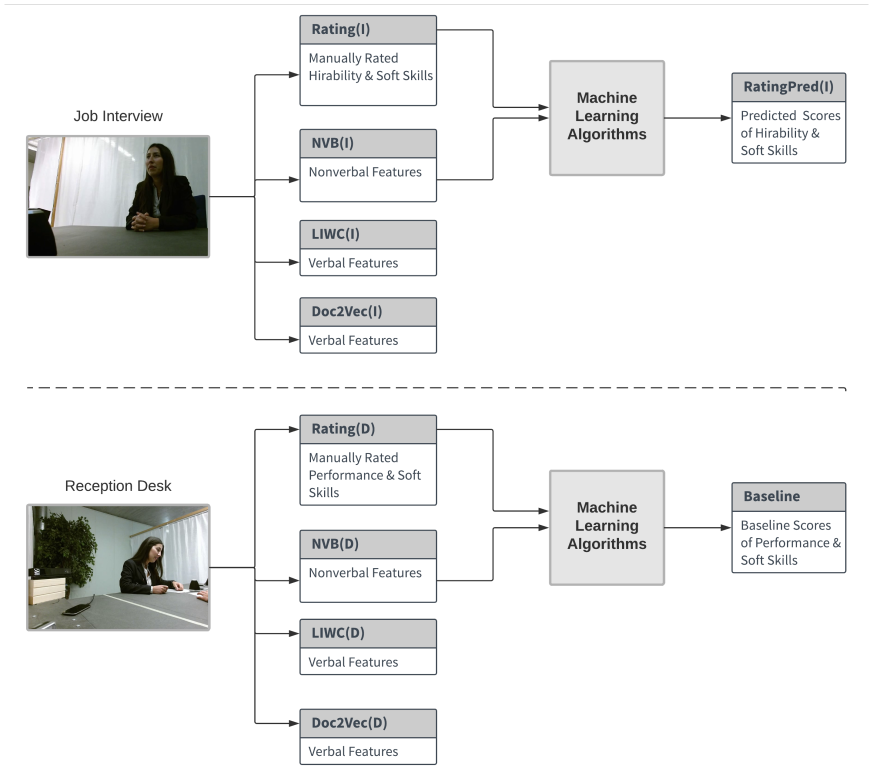


Fig. 4. A visual summary of the cues used in our experiments and how they were obtained.

5 INFERENCE FRAMEWORK AND EXPERIMENTAL PROTOCOL

In this section, we outline the inference framework and experimental protocol. The various input components for our experiments and their source are visualized in Figure 4. As a first step towards answering RQ1 and RQ2, we perform a correlation analysis between the variables of interest. We report only the correlation values which are significant with $p < 0.05$. We then define a regression task in which perceived performance and soft skills at the reception desk is inferred from nonverbal cues, ratings from job interviews, nonverbal cues from reception desk, linguistic content from both situations, and various combinations of these features.

For the regression tasks, we follow a standard machine learning protocol. First, the data was pre-processed using a person-independent Z-score normalization to transform data into unity variance and zero mean. Then, two unsupervised dimensionality reduction techniques were evaluated:

(1) Low p-value features(p-val): In this method, features which were significantly correlated ($p < 0.05$) only were selected. This is based on the assumption that important information is encoded in significantly correlated features.
(2) Principal Component Analysis (PCA): This method projects the features into a lower dimension orthogonal space [63].

The performance of these dimensionality reduction did not improve performance over the use of the original features and hence, their results are not reported here. Two regression techniques (Support Vector Machines

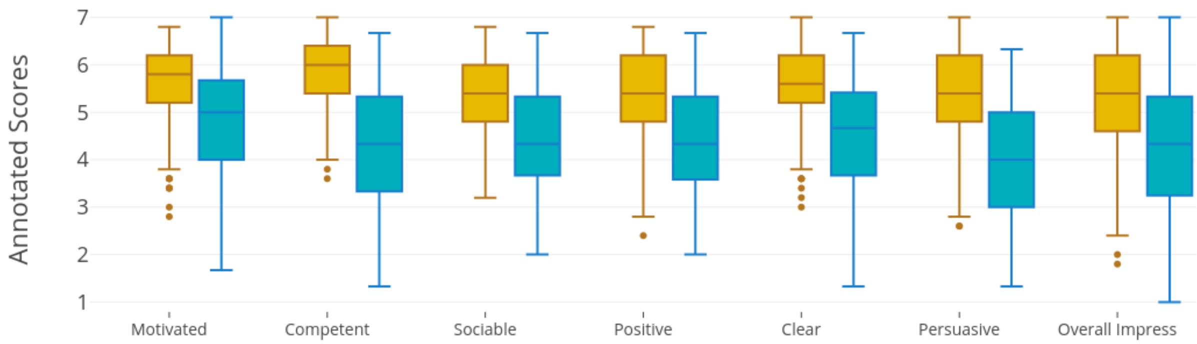


Fig. 5. Box-plot showing the distribution of annotated scores for each of the variable of interest. Here we observe that mean scores for interview (yellow) is greater than mean scores for reception desk (blue).

regression (SVM-R) [21] and Random Forest regression (RF) [14]) were evaluated using the Caret package [46] for R implementation. These algorithms were selected to understand the contributions of each component in inferring perceived performance and soft skills in the reception desk situation. The hyper parameters of the machine learning algorithms were optimized using 10-fold inner cross-validation (CV), while the performance was assessed using the 100 independent runs of leave-one-video-out CV. The performance of these regression techniques was evaluated by employing coefficient of determination (R^2). We use the R^2 values reported in our previous work [56], obtained using nonverbal behavioral cues only, as the baseline for comparing results.

6 RESULTS AND DISCUSSION

We now present the results and discussion corresponding to each of the three RQs we originally posed.

6.1 RQ1: Perceived Variables in Interview and Reception Desk Situations

We begin by computing the descriptive statistics of the perceived variables of both situations, presented as box plots in Figure 5. We observe that the mean ratings for all perceived variables in the reception desk situation are lower than the corresponding ratings in the job interviews, indicating that all variables were more favorably perceived in the interview than at the reception desk. We hypothesize that this is due to the reception desk interactions occurring in a more challenging situation (the client is unhappy and not easy to persuade) while the interview interactions occur under positive tone. This hypothesis has backing in psychology, which suggests that positive evaluations tend to occur under positive mood [32, 50]. To verify if the difference in perceived

Table 2. Pearson's correlation between perceived variables from interview (I) and reception desk (D) situations ($N = 169$). All of them are significant with $p < 0.001$

I.Motivated	I.Competent	I.Positive	I.Sociable	I.Clear	I.Persuasive	I.Hirability
0.49	0.41	0.44	0.49	0.30	0.40	0.45
D.Motivated	D.Competent	D.Positive	D.Sociable	D.Clear	D.Persuasive	D.Performance

Table 3. List of predictors used in regression experiments obtained from job interview (I) and reception desk (D) interactions.

Abbreviation	Details
NVB (I)	Nonverbal behavior extracted from interviews
Ratings (I)	Manually rated hirability and soft skills from interviews
RatingPred(I)	Automatically predicted scores of hirability and soft skills from interviews
LIWC(I)	LIWC features extracted from manual transcriptions of interviews
Doc2Vec(I)	Doc2Vec features extracted from manual transcriptions of interviews
NVB (D)	Nonverbal behavior extracted from desk
Ratings (D)	Manually rated performance and soft skills from desk
LIWC(D)	LIWC features extracted from manual transcriptions of desk
Doc2Vec(D)	Doc2Vec features extracted from manual transcriptions of desk
LIWC(D + I)	Combined LIWC features extracted from manual transcriptions of interviews and desk

variables across the settings was significant, we conducted a test of means for each variable. As the population of participants was the same across both the settings, we used a paired Student T-test ($N = 169$). The test of means refuted the *Null* hypothesis ($p < 0.001$), indicating that the differences in mean perceived variable scores during the job interview were significantly higher than the mean perceived variable scores in the desk situation.

Correlation Analysis: We conducted a Pearson's correlation analysis on the perceived variables from the two situations. Results of this analysis are presented in Table 2. We observe that all perceived variables are positively correlated to each other ($p < 0.001$ in all cases). *Sociable* and *Motivated* in the two situations have the highest correlation ($r = 0.49$), while *Clear* has the lowest ($r = 0.30$). An interesting observation is the correlation between perceived performance and perceived hirability ($r = 0.45$). This seems to suggest that participants who were perceived as more hirable during their interviews were to some degree perceived to perform better on the job.

Inference Task: We then investigated the ability of the perceived variables from job interviews in inferring perceived performance and soft skills at the job situation as a regression task. The baseline for this work is the R^2 obtained using nonverbal cues to infer perceived variables, specifically *Performance* ($R^2 = 0.30$) reported in our previous work [56].

Table 3 summarizes the various predictors used in all our inference experiments. Towards answering **RQ1**, we define four experiments, labeled Exp1a-Exp1d Table 4, to test different conditions involving perceived variable scores (visualized in Figure 4). In *Exp1a*, we use the perceived scores from job interviews as predictors of perceived performance and soft skills. We observe that the best performance of these perceived variables (using SVM-R) was slightly lower than *Baseline* with $R^2 \in [0.18, 0.27]$. The best performance was observed for *Sociable* ($R^2 = 0.27$) and lowest for *Clear* ($R^2 = 0.18$). This set of predictors produces $R^2 = 0.25$ for *Performance*. These results can be explained by the correlations between the perceived variable in the two situations. Our results show that perceived performance and soft skills on the job can be inferred to some extent by just the perceived hirability and soft skills scores during interviews.

To further understand this connection, we conducted another regression task using automatically predicted scores from the job interviews instead of manually generated scores, to study a situation where fully automatic assessment at the interview could be used to make inference at the job (*Exp1b*). The predicted scores (RatingPred(I)) were obtained by using nonverbal cues displayed during job interviews as predictors in a regression task with random forest (RF). This method has been shown to result in $R^2 = 0.32$ [55, 58]. A paired test of means accepted the null hypothesis, indicating that the means of predicted scores and manual scores were not statistically significant. However, the use of these predicted scores (RatingPred(I)) for regression at the job showed a lower

Table 4. Summary of experiments and the best regression performance (R^2) achieved. All results are significant with $p < 0.05$.

Experiment	Predictors	Model	motiv	compe	posit	socia	clear	persu	perfo
Baseline	NVB (D)	RF	0.30	0.29	0.32	0.33	0.22	0.32	0.30
Exp1a	Ratings (I)	SVM-R	0.21	0.23	0.26	0.27	0.18	0.24	0.25
Exp1b	RatingPred(I)	SVM-R	0.27	0.26	0.28	0.26	0.16	0.22	0.21
Exp1c	Ratings (I) + NVB (D)	RF	0.34	0.37	0.36	0.39	0.24	0.37	0.40
Exp1d	RatingPred(I) + NVB(D)	RF	0.36	0.32	0.32	0.34	0.23	0.30	0.31

performance for four of the variables like *Clear, Persuasive* and *Performance* ($R^2 \in [0.16, 0.28]$). Even though the performance was lower, this result shows a first step towards using automatically inferred scores of job interviews to infer perceived performance and soft skills at the job.

The best performing model was obtained in *Exp1c*, where we studied the effect of combining nonverbal cues displayed at the desk and scores of perceived variables from interviews (using RF). We obtain $R^2 = 0.40$ for *Performance*, compared to a baseline of $R^2 = 0.30$. An improved inference performance is also observed for other variables with *Sociable* ($R^2 = 0.39$), *Competent, Persuasive* ($R^2 = 0.37$), *Positive* ($R^2 = 0.36$), with the lowest performance for *Clear* ($R^2 = 0.24$). To complete the experiments, we infer the impressions at desk using automatically predicted scores from the interview (*Exp1d*) in addition to nonverbal cues extracted from reception desk. The results indicate that this fully automated condition brings about marginal improvement over the baseline like *Performance* (from 0.30 to 0.31), *Clear* (from 0.22 to 0.23), *Sociable* (from 0.33 to 0.34), and *Competent* (from 0.29 to 0.32).

As the next step, to understand the contributions of features to infer perceived performance, we list the top 20 variables used by the RF algorithm (Table 5). This list was obtained by using the *var.Imp* function in CARET, which returns the variables and their measure of importance (scaled to 100). We observe that this list of top variables includes scores of perceived variables from job interviews, nonverbal cues from both the participants and the clients. Specifically, we observed that *Hirability* and *Persuasive* scores rated at the interview were marked as two of the seven most important variables by RF. Similarly, participant cues found to contribute include speaking time, turn duration (mean and max), head nods (mean and duration), voice energy modulation (upper and lower quartile), and visual back-channeling (duration). An interesting observation is that client nonverbal cues like speaking energy, voice energy modulation, and spectral entropy also contribute to inference performance.

To summarize, in this subsection we investigated the question: How are perceived variables in the job interview connected with perceived variables on the job situation? Our main results are: (1) Scores of perceived variables from job interviews and perceived variables in reception desk are moderately correlated. (2) The perceived variables scores at reception desk can be inferred to some extent ($R^2 \in [0.21, 0.25]$) from perceived variables in job interviews, both manual rated and automatically inferred. (3) The fusion of the perceived variable scores from job interview and nonverbal cues extracted from the desk improves inference of perceived variables at the desk, with a best performance of $R^2 = 0.40$. Our results indicate that the impressions made during job interviews add information to the nonverbal behavior during the desk situation.

6.2 RQ2: NVB in Interviews and Perceived Performance at the Desk

In this subsection, we investigate the links between automatically extracted nonverbal behavior of candidates during job interviews and the perception of performance on the job. We first present a correlation analysis and then the inference task.

Table 5. Top 20 variable importance in the RF for *Exp1c*. All measures of importance indicated in the *Rank* column are scaled to have a maximum value of 100.

Cues	Rank	Cues	Rank
Participant cues			
Speaking time	100.00	Total number of head nods	64.41
Speaking ratio	80.01	Upper quartile of change in speaking energy	62.98
Mean duration of nodding while speaking	74.25	Std of turn duration	62.22
Number of nods while speaking	73.22	Lower quartile of change in speaking energy	62.08
Mean turn duration	72.83	Max duration visual back-channeling	59.18
Max turn duration	67.15		
Interview Ratings			
Persuasive	87.67	Motivated	58.44
Hirability rating (Interview)	69.56	Communicative rating (Interview)	57.18
Enthusiastic	61.05		
Client cues			
Lower quartile speaking energy	67.78	Min spectral entropy	59.45
Upper quartile of change in speaking energy	60.15	Max speaking energy	57.18

Correlation Analysis: As a first step, we conduct a Pearson's correlation analysis between nonverbal cues extracted during the job interviews and perceived variables at the desk situation. The results that show weak to moderate trends are presented in Table 6. We observe that participants who spoke for longer duration, with less silence, and had greater speaking energy modulation during the job interview, were perceived to perform better at the reception desk. Also, participants who nodded more, for greater duration, displayed greater number of visual back-channeling, and nodded more while speaking were rated as better performing during the reception desk.

These results are supported by literature in psychology [24, 53]. In [53], Motowidlo et al. using a dataset of simulated job interviews of 40 managers reported correlations of $r = 0.32$ between visual features and performance ratings, $r = 0.33$ between aural features and performance ratings, and $r = 0.36$ between combined aural and visual features and performance ratings. In that work, supervisors' ratings were considered as performance ratings. Similar results were reported in another work by DeGroot et al. [24]. Using videotaped interviews of 110 managers in a news-publishing company, it was reported that vocal cues correlated with performance ratings with $r = 0.20$ ($p < 0.05$). That work also found low correlations of $r = 0.14$ ($p < 0.05$) between performance ratings and composite visual cues (like physical attractiveness, smiling, gaze, hand movement, and body orientation).

We then compute Pearson's correlation between nonverbal behavioral cues in the two situations and perceived variable scores (Table 7). Specifically, we compute (a) correlation between nonverbal cues extracted from interviews (NVB(I)) and perceived hirability (Rating(I)) (b) correlation between nonverbal cues extracted from reception desk (NVB(D)) and perceived performance (Rating(D)). We observe that participants who displayed specific behavioral patterns had a weak-to-moderate trend to be rated high in both situations. Specifically, participants who spoke longer, louder, had fewer silence events were perceived as more hirable during the interview and also perceived as better performing on the job. Similarly, participants were perceived more positively when they moved more, nodded more and for longer time. This suggests that positive impressions could be related to similar behavioral cues in the two situations. Our results are in accordance with existing literature for interviews, where participants speaking for longer, with fewer silence, nodded more were rated as more hirable [23, 55, 58].

Table 6. Pearson's correlation between perceived variables of desk and nonverbal cues displayed during job interviews ($N = 169$). All features are significant $^{***}p < 0.001$,$^{**}p < 0.01$, $^{*}p < 0.05$

	Ratings(D)						
NVB(I)	motiv	compe	posit	socia	clear	persu	peImp
Speaking Activity Features							
Num speaking turns	-0.23**	-0.24**	-0.18*	-0.15	-0.26**	-0.28**	-0.21*
Mean turn duration	0.33***	0.24**	0.21*	0.17*	0.23**	0.25**	0.24**
Number of silence events	-0.36***	-0.30***	-0.31***	-0.32***	-0.27**	-0.35***	-0.30***
Silence Ratio	-0.36***	-0.31***	-0.31***	-0.33***	-0.29***	-0.34***	-0.27**
Prosodic Features							
Lower quartile speaking energy	0.27**	0.22**	0.31***	0.35***	0.21*	0.22*	0.23**
Max speaking energy change	0.24**	0.22**	0.22*	0.28***	0.21*	0.18*	0.22*
Visual Features							
Total num head nod	0.41***	0.36***	0.38***	0.45***	0.33***	0.32***	0.33***
Total duration head nod	0.39***	0.38***	0.37***	0.43***	0.35***	0.36***	0.35***
Num of nod speak	0.37***	0.29***	0.32***	0.39***	0.26**	0.25**	0.28**
Mean duration of nod speak	0.35***	0.28**	0.22*	0.31***	0.23**	0.26**	0.27**
Std duration nod speak	0.34***	0.28***	0.24**	0.33***	0.22**	0.26**	0.26**
Max duration nod speak	0.40***	0.33***	0.30***	0.40***	0.26**	0.30***	0.30***
Num visual BC	0.31***	0.33***	0.31***	0.36***	0.31***	0.32***	0.29***
Mean duration visual BC	0.23**	0.30***	0.27**	0.27**	0.31***	0.31***	0.28**
Std duration visual BC	0.23**	0.28**	0.25**	0.29***	0.26**	0.26**	0.27**
Max duration visual BC	0.28**	0.30***	0.28***	0.30***	0.28**	0.29***	0.29***

Inference Task: We then use regression to infer the perceived performance and soft skills at the reception desk from automatically extracted nonverbal cues from the interviews. A total of four experiments labeled Exp2a-Exp2d, were conducted using the various components illustrated in Figure 4 and the results are presented in Table 8. As a first step (*Exp2a*), we use all the nonverbal cues displayed during interviews to infer perceived performance and soft skills at the desk. We observe that these nonverbal cues overall have low predictive power with $R^2 \in [0.12, 0.30]$. The best performance is achieved for *Sociable* ($R^2 = 0.30$) and lowest for *Clear* ($R^2 = 0.12$). For *Performance*, this model achieved $R^2 = 0.17$. Though these results are lower then the baseline, they indicate a weak connection between behavioral cues from job interview and perceived variables on the job.

In the next step, we fuse the nonverbal cues extracted from the two situations and use them as predictors. The idea is to investigate the effect of extra behavioral information on the inference performance. In *Exp2b*, we observed that inference of some variables improved as compared to the baseline. Specifically, there is improvement for *Motivated* (from 0.30 to 0.34), *Sociable* (from 0.33 to 0.36) and *Competent* (from 0.29 to 0.30) while for *Positive* and *Persuasive* the performance decreased slightly. Fusion of nonverbal cues from both situations had no effect on inference of *Performance* at desk.

We then combine nonverbal cues and perceived scores from interviews to infer perceived variables at reception desk (*Exp2c*). The performance varies with $R^2 \in [0.26, 0.35]$ with best performances for *Positive* and *Sociable* ($R^2 = 0.35$) followed by *Motivated* ($R^2 = 0.32$), *Competent, Persuasive* ($R^2 = 0.28$) and *Performance* ($R^2 = 0.26$). This is the best result achieved using all information available from the job interviews and is comparable to

Table 7. Selected Pearson's correlation coefficient for perceived hirability (Rating(I)) and perceived performance (Rating(D)) across the two situations ($N = 169$). $^{**}p < 0.01$, $^{*}p < 0.05$

Nonverbal Cues	Perceived Hirability	Perceived Performance
Acoustic Features		
Avg Turn duration	0.39^{**}	0.40^{**}
Speaking Ratio	0.21^{**}	0.43^{**}
Num Silent Events	-0.43^{**}	-0.22^{**}
Speaking Energy (Q25)	0.29^{**}	0.18^{*}
Speaking Energy Derivative (Q25)	-0.27^{**}	-0.20^{*}
Visual Features		
Mean WMEI	0.18^{*}	0.26^{*}
Max WMEI	0.16^{*}	0.30^{*}
Total Head Nod	0.25^{**}	0.37^{**}
Num of Nods while speaking	0.26^{**}	0.45^{**}
Max duration of Nods while speaking	0.25^{**}	0.39^{**}

Table 8. Summary of experiments and best regression performance (R^2) of desk perceived variables achieved. All results are significant with $p < 0.05$.

Experiment	Predictors	Best Model	motiv	compe	posit	socia	clear	persu	perfo
Baseline	NVB (D)	RF	0.30	0.29	0.32	0.33	0.22	0.32	0.30
Exp2a	NVB (I)	RF	0.24	0.18	0.17	0.30	0.16	0.12	0.17
Exp2b	NVB (I) + NVB (D)	RF	0.34	0.30	0.27	0.36	0.22	0.29	0.30
Exp2c	NVB (I) + Ratings (I)	SVM-R	0.32	0.28	0.35	0.35	0.25	0.29	0.26
Exp2d	NVB (I) + Ratings (I) NVB (D)	RF	0.33	0.28	0.32	0.39	0.27	0.32	0.32

Baseline, importantly, without seeing *any* data at the job. As a final experiment (*Exp2d*), we fused perceived scores at job interviews and all the nonverbal cues extracted from both situations and use them as predictors. We observe a slightly improved performance compared to the *Baseline* with the highest variance explained for *Sociable* ($R^2 = 0.39$), followed by *Positive, Persuasive*, and *Performance* ($R^2 = 0.32$).

To summarize, there are two main findings in this subsection: (1) Some nonverbal cues like speaking and turn duration, head nods displayed during job interviews are weakly-to-moderately correlated to perceived performance and soft skills at the reception desk. This result could have implications for behavioral training systems where focus can be on specific behaviors for multiple situations. (2) We observed that nonverbal cues extracted from job interviews have weak inference ability ($R^2 = 0.17$). Importantly, this performance improves ($R^2 = 0.26$) when these nonverbal features are augmented with perceived scores from job interviews. These results suggest that for some soft skills displayed in the actual job, it is useful to use behavior and impressions from the interview situation.

6.3 RQ3: Linguistic Content and Perceived Performance

To address *RQ3*, we conducted nine experiments Exp3a-Exp3i, with different linguistic features extracted. Here again the *Baseline* is the performance obtained in inferring perceived performance using nonverbal cues extracted

Table 9. Summary of experiments with linguistic content and the best inference performance achieved. All results are significant with $p < 0.05$.

Experiment	Predictors	Model	motiv	compe	posit	socia	clear	persu	perfo
Baseline	NVB (D)	RF	0.30	0.29	0.32	0.33	0.22	0.32	0.30
Exp3a	LIWC(D)	RF	0.09	0.22	0.19	0.14	0.24	0.17	0.18
Exp3b	LIWC(I)	RF	0.04	0.04	0.02	0.01	0.13	0.02	0.07
Exp3c	LIWC(D + I)	RF	0.15	0.22	0.19	0.14	0.25	0.17	0.18
Exp3d	Doc2Vec(D)	SVM-R	0.08	0.09	0.07	0.05	0.06	0.08	0.10
Exp3e	Doc2Vec(I)	SVM-R	0.16	0.18	0.18	0.15	0.10	0.09	0.16
Exp3f	LIWC(D) + Doc2Vec(D)	SVM-R	0.17	0.22	0.16	0.19	0.19	0.20	0.25
Exp3g	LIWC(I) + Doc2Vec(I)	SVM-R	0.24	0.24	0.17	0.17	0.18	0.16	0.26
Exp3h	LIWC(D) + NVB(D) Doc2Vec(D)	RF	0.25	0.29	0.27	0.32	0.27	0.29	0.26
Exp3i	LIWC(I) + NVB(I) Doc2Vec(I)	RF	0.28	0.18	0.16	0.29	0.11	0.15	0.20

from the reception desk ($R^2 = 0.30$). The input for these experiments is illustrated in Figure 4 and the results are tabulated in Table 9.

LIWC:. First, we use LIWC to extract lexical cues from the reception desk transcribed data (LIWC(D)) and use them to infer perceived performance and soft skills (*Exp3a*). LIWC features show lower performance than the *Baseline* with $R^2 \in [0.09, 0.24]$ for all variables. The best performance was for *Clear* ($R^2 = 0.24$) followed by *Competent* ($R^2 = 0.22$), and *Motivated* ($R^2 = 0.09$) being the worst. Linguistic content of the desk results in $R^2 = 0.18$ for *Performance*. Note that this is better than results reported in investigations of linguistic content and *Overall Impression* in job interviews ($R^2 = 0.11$) in the literature [54, 57].

In a second step (*Exp3b*), using LIWC features extracted from job interviews (LIWC(I)), we find that the performance of linguistic content in inferring perceived performance and soft skills is very low, with $R^2 < 0.1$ for almost all variables (except *Clear*, $R^2 = 0.13$). We then combine the LIWC features from both settings (LIWC(D+I)) to infer impressions of performance and skills (*Exp3c*). We find no improvement except for *Motivated*.

Doc2Vec: We then investigate the potential of Doc2Vec with features extracted using the reception desk, Doc2Vec(D) in *Exp3d*. Interestingly, the performance of the Doc2Vec(D) is lower than the LIWC(D) features with $R^2 \in [0.05, 0.10]$, with $R^2 = 0.10$ for *Performance*. This is in contrast to results reported in the literature for job interviews [17]. In that work, the authors using Doc2Vec features to infer *Hirabilty* scores from 36 job interviews, and reported a correlation $r = 0.41$ between manual and automatic hirability. Converting r to R^2 for comparison, this work achieved $R^2 = 0.16$. We believe that the low performance we obtain could be due to the relatively short duration of the reception desk interactions, which has an average of 354.1 words for all turns taken by the participant. The authors of [17] have not reported the corpus size used in their work so a direct comparison is not possible. As a next step, we use the Doc2Vec(I) features consisting of Doc2Vec features extracted from the job interviews (*Exp3e*). Interestingly, the performance with this feature set was better than the one achieved by Doc2Vec(D), with $R^2 \in [0.09, 0.18]$. The use of these features produces $R^2 = 0.16$ for *Performance* and is similar in range to those reported by Chen et al. [17]. We believe this improvement in performance might be due to the larger duration of the job interviews. This corpus is more than twice as long as the reception desk corpus, and contains an average of 813 words.

Fusion of LIWC & Doc2Vec: As a next step, we combine the two linguistic features. In *Exp3f*, we combine LIWC(D) and Doc2Vec(D) as predictors. We observe that the inference performance is better than each of the features individually, with $R^2 = 0.25$ for *Performance* as compared to $R^2 = 0.10$ and $R^2 = 0.18$ for Doc2Vec(D) and LIWC(D), respectively. The improvement is observed for all the perceived variables of the reception desk. Similarly, the fusion of linguistic features from the interviews (*Exp3g*) also leads to an improved performance for all the perceived variables as compare to each of individual features. With the fused feature set, we observe $R^2 = 0.26$ for *Performance* is explained. This is the best performance achieved using the linguistic features.

Fusion of Linguistic and Nonverbal Features: In the final step, we use a fusion of nonverbal cues and linguistic features from the reception desk situation as predictors (*Exp3h*). Except for one variable (*Clear*, $R^2 = 0.27$) these results are not better than the *Baseline* performance. The same is the case when we combine nonverbal and linguistic features from the interview situation (*Exp3i*), which do not improve over the baseline (NVB(D)). To understand this result, we listed the top 20 variables used by the RF algorithm using the *var.Imp* function in CARET. We do not report them here as this list did not contain any verbal features in top 20 and hence not very helpful in understanding the impact of linguistic features on perceived variables.

To summarize, the main findings of this section are (1) LIWC features outperforms the Doc2Vec features using the reception desk data, with the best performance being always worse than using nonverbal cues. (2) Interestingly, the Doc2Vec features from job interviews perform comparable to LIWC features from desk. (3) The fusion of LIWC and Doc2Vec features from both situation results in improved inference, with *Exp3f* giving $R^2 = 0.25$, while *Exp3g* gives $R^2 = 0.26$. Overall, linguistic features can, moderately and to a lesser degree than nonverbal behavior, be useful to infer perceived performance and soft skills.

6.4 Qualitative Study

To understand the implications of this work for real-world situations in hospitality, we conducted a small qualitative study. The study consisted of two sets of questionnaires consisting of five questions (Appendix A). One set of questions was sent to ten selected participants of the study, while the other was sent to two directors of the hospitality school where the dataset was collected. Of the 12 people contacted, we received responses from four people, two participants (henceforth called student A and B) and two directors (henceforth called director C and D).

Specifically, we asked the hospitality student recipients of the questionnaire about their experience during the interview role play and its relation to the real world. Student A said that “It was a good experience to realize how stressful an interview could be. I was very happy to do it because a few months after I had to do a ‘true’ interview and they asked me similar questions! I was feeling prepared because I knew how to deal with it. Like if I prepared an exam.” Student B, replying to a question about the experience during the reception desk role play and its relation to the real world, said “It is a very common situation that is faced in reception so it was a very appropriate exercise linked to our line of work. Being able to handle dissatisfied guests or situations under pressure is good practice for us to learn, how to foresee situations or be proactive within our line of work for the future”. Both students said they did not know that there might exist a link between interview ratings and perceived performance on the job. Similarly, both students felt that the use of technology in specific stressful situations (like angry clients at a reception desk) can help them improve their nonverbal behavior. Specifically, student A said, “It can give us a perspective that we may not have noticed before or change our opinion on a certain behavior”.

The two directors of the hospitality school too felt that the role-playing during data collection had connections to the real world situations in hospitality. Specifically, director C expressed the opinion that role-playing could help young hospitality professionals as "One can play out different scenarios about guest contact, without actually throwing the person in at the deep end. In other words, one can practice without real guests' difficult situations which in turn will assist the young professional once he/she encounters them." Director D was on the opinion that role-playing helps students gain insights on the nuances of what the job entails. He, however, cautioned that "in the real world, even if our students have the required skills and personality to be hired in a position, the most challenging issues will be to adapt to a new people and environments as well as to adopt new style of work."

Furthermore, the two directors were enthusiastic about the role of technology in improving young professionals' behavior on the job. Director C responded by saying that "technology can be used as a mirror of actions and to effectively communicate desired behavior patterns that can serve as a role-model for people on the job". Director D felt that by using technology "students and young professionals could train and improve their behaviors and speeches when faced with different types of clients: introverts, extroverts, violent, sly".

In summary, both the participants and the directors of the school felt that role-playing of job-related situations helped hospitality students be prepared for stressful situations like facing an interview or a difficult client at the reception desk. They were also enthusiastic about using ubiquitous computing systems to capture and analyze behavioral cues as they felt they can help achieve better behavioral awareness during professional interactions in customer-facing jobs.

7 IMPLICATIONS

We believe that the insights from our work have implications for hospitality and other customer-facing domains where interpersonal communication and soft skills are critical. Here, we discuss implications of our work in the domains of hospitality, psychology, and ubiquitous computing.

In the hospitality industry, there is great emphasis on soft skills and interpersonal communication as they are considered critical to business [29, 80]. Our work contributes to this domain by showing connections between the automatically extracted nonverbal behavior of potential employees (displayed during a job interview), and their perceived performance on the job. As the objective of job interviews is to select the best candidate for a given position, our work provides important inferences for human resource teams and hiring managers in the domain of hospitality. Our research also shows the feasibility of utilizing an automatic framework using both verbal and nonverbal behavior for assessing candidates for customer-facing roles. Given that we have identified some of the most important nonverbal behaviors in the job interview that are moderately connected to future job performance (speaking longer, louder, with fewer silences, gesturing more while speaking and nodding for longer periods of time), training in hospitality and other service-related fields might put an emphasis on students learning these behaviors and maintaining them also under stressful conditions (e.g., dissatisfied clients).

In psychology, our work is a step in the direction of understanding human behavior in multiple situations by integrating ubiquitous computing and social psychology. Our research shows the importance of face-to-face job interviews for their predictive value for perceived performance in customer-facing jobs. Recruiters underscore the importance of a personal meeting with a job applicant and often talk about the importance of "feeling" the applicant. This "feeling" might refer to observing the applicant's nonverbal behavior and one's own nonverbal reactions to it. This is in line with our findings indicating that some of the applicant's and the interviewer's nonverbal cues have predictive power on the perceived performance on the job (reception desk) [24, 53].

Finally, in ubiquitous computing, our work has implications for developing behavioral training systems. Specifically, the observation that the same behavioral cues are positively linked to first impressions across settings encourages the development of behavioral awareness systems that focus on specific cues. Such systems could be helpful for individuals who aspire to improve the nonverbal behavior they convey [28, 37]. Such systems are also potentially important to socially challenged individuals to express and/or perceive nonverbal communication. Overall, understanding differences in behavior across situations and the information they convey is important to build ubiquitous computational devices capable of sensing and responding unobtrusively [61, 84].

8 CONCLUSION

This paper describes our investigation into human behavior (verbal and nonverbal) and formation of impressions across multiple situations using ubiquitous sensing and multimodal analysis. Specifically, we investigated the connections between verbal content, displayed nonverbal behavior, and perceived variables under two different situations in the context of hospitality. Towards this, we used a data corpus of 338 interactions, recorded in multiple modalities and role-played by hospitality students in two settings; job interview and reception desk. A number of nonverbal behavioral cues were automatically extracted. Further, the interview and desk interactions were manually transcribed, and then verbal cues were extracted from these transcriptions.

We posed three research questions (RQs) summarized here:

RQ1 examined the connections between perceptions of candidates in job interviews and perceptions of the same person on the job. The four main findings were: (1) mean scores of perceived variables were higher in the job interviews than the corresponding ratings on the job. This implies that participants were perceived more favorable during the interviews. (2) We observed that perceived variables from job interviews were weakly to moderately positively, correlated to perceived variables in the reception desk situation. (3) We found that perceived variables on the job can be inferred, to some extent, from manually rated perceived variables ($R^2 = 0.25$) and automatically inferred scores ($R^2 = 0.21$) in job interviews. (4) The fusion of automatically extracted nonverbal cues from the desk situation with the perceived variable scores from interviews improved inference of perceived variables on the job, and corresponding best performance ($R^2 = 0.40$).

RQ2 examined the link between automatically extracted nonverbal behavior of candidates during job interviews and the perception of performance on the job. There were two main findings: (1) Some nonverbal cues displayed during job interviews were weakly to moderately correlated to perceived performance and soft skills at the reception desk situation. (2) Using these nonverbal cues as predictors in an inference task had a moderate performance with $R^2 = 0.17$ for perceived performance. Augmenting these nonverbal features with perceived variable scores in job interviews, the performance improved with $R^2 = 0.26$. Our result indicates a moderate feasibility to use nonverbal cues displayed during job interviews in inferring perceived performance and soft skills in the reception desk setting.

RQ3 studied the connections between candidates' choice of words in the two interactions and the perception of performance on the job. This results revealed some feasibility of using linguistic features to infer perceived variables on the job, although their performance is lower for all the perceived variables than the baseline. The three main findings were: (1) LIWC features extracted from reception desk outperformed the Doc2Vec features computed from the same situation in the inference of perceived variables on the job. (2) The Doc2Vec features extracted from job interviews performed comparably to LIWC features extracted from the reception desk situation. (3) Fusing the LIWC and Doc2Vec features from the desk situation improved inference performance, with $R^2 = 0.25$ achieved for Performance.

In summary, we believe this work is a step in the direction of understanding human behavior in multiple situations combining ubiquitous computing and social psychology. Our results could have implications for the hospitality industry but also for other customer service jobs, and for individuals who are socially challenged and unable to express and/or perceive nonverbal communication. Hence, understanding human behavior and the social signals they convey across multiple situations is an important step towards building ubiquitous devices capable of sensing and responding unobtrusively.

In the future, we plan to take advantage of the advances in automatic speech recognition (ASR) to understand its impact on automatic inference of perceived variables and to extend the work done to address **RQ3**. We also plan to investigate in depth the cases with inconsistencies in nonverbal behavior across the two situations, with an aim to better understand and model such cases. Specifically, we could study the behavioral differences between participants in the top and bottom quartiles of ratings for all variables.

ACKNOWLEDGMENTS

This work was supported by the Sinergia program of the Swiss National Science Foundation (SNSF) through the UBImpressed project. We would like to thank Denise Frauendorfer, Laurent Son Nguyen and our research assistants for their help with data collection, Yu Yu and Jean-Marc Odobez for providing the head nod features, and the students of Vatel Switzerland for their participation.

APPENDIX

A QUESTIONS IN QUALITATIVE STUDY

A.1 Questionnaire for Students

(1) How does the role playing during the interview experiment relate to your real-world experience?
(2) How does the role playing during the front-desk experiment relate to your real-world experience?
(3) Did you know that there are connections between how well you did in the interview and how well you do on the job.
(4) How in your opinion can technology help you improve your behavior during job interviews?
(5) How in your opinion can technology help you improve your behavior on the job?

A.2 Questionnaire for Directors

(1) In your experience, how do you use the information from job interviews to forecast on-the-job performance of a young hospitality employee?
(2) How, in your opinion, can technology help young hospitality professionals to improve their behavior during job interviews?
(3) How, in your opinion, can technology help young hospitality professionals to improve their behavior on the job (like the front desk)?
(4) What is the value of role-playing job interviews in relation to the real-world experience of job interviews for young hospitality professionals?
(5) What is the value of role playing in relation to the real-world experience of front desk interactions for young hospitality professionals?

REFERENCES

[1] Alexander T Adams, Jean Costa, Malte F Jung, and Tanzeem Choudhury. 2015. Mindless computing: designing technologies to subtly influence behavior. In *Proc. ACM UBIComp*. ACM, 719–730.
[2] Nalini Ambady, Mary Anne Krabbenhoft, and Daniel Hogan. 2006. The 30-sec sale: Using thin-slice judgments to evaluate sales effectiveness. *J. Consumer Psychology* 16, 1 (2006), 4–13.
[3] Nalini Ambady and Robert Rosenthal. 1992. Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin* 111, 2 (1992).
[4] Nalini Ambady and Robert Rosenthal. 1993. Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *J. Personality and Social Psychology* 64, 3 (1993), 431.
[5] Nalini Ambady and John Joseph Skowronski. 2008. *First impressions*. Guilford Press.
[6] Neil Anderson and Viv Shackleton. 1990. Decision making in the graduate selection interview: A field study. *J. Occupational Psychology* 63, 1 (1990), 63–76.

[7] 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. ACM ICMI*.

[8] Marjorie A Bayes. 1972. Behavioral cues of interpersonal warmth. *J. Consulting and Clinical Psychology* 39, 2 (1972), 333.

[9] Joan-Isaac Biel and Daniel Gatica-Perez. 2013. The YouTube lens: Crowdsourced personality impressions and audiovisual analysis of vlogs. *IEEE Trans. on Multimedia* 15, 1 (2013).

[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, 119–126.

[11] Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural language processing with Python: analyzing text with the natural language toolkit*. O'Reilly Media, Inc.

[12] Silvia Bonaccio, Jane O'Reilly, Sharon L O'Sullivan, and François Chiocchio. 2016. Nonverbal behavior and communication in the workplace: A review and an agenda for research. *J. Management* 42, 5 (2016), 1044–1074.

[13] Joyce E Bono and Remus Ilies. 2006. Charisma, positive emotions and mood contagion. *The Leadership Quarterly* 17, 4 (2006), 317–334.

[14] Leo Breiman. 2001. Random forests. *Machine learning* 45, 1 (2001), 5–32.

[15] Judee K Burgoon, Thomas Birk, and Michael Pfau. 1990. Nonverbal behaviors, persuasion, and credibility. *Human Communication Research* 17, 1 (1990), 140–169.

[16] Ron Caneel. 2005. *Social signaling in decision making*. Ph.D. Dissertation. Massachusetts Institute of Technology.

[17] 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, 161–168.

[18] Yiqiang Chen, Yu Yu, and Jean-Marc Odobez. 2015. Head Nod Detection from a Full 3D Model. In *Proc. IEEE ICCV Workshops*.

[19] Cindy Chung and James W Pennebaker. 2007. The psychological functions of function words. *Social Communication* (2007), 343–359.

[20] Microsoft Cognitive Services. [n. d.]. Azure Emotion API. https://azure.microsoft.com/en-us/services/cognitive-services/emotion/

[21] Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Machine learning* 20, 3 (1995), 273–297.

[22] Jared R Curhan and Alex Pentland. 2007. Thin slices of negotiation: predicting outcomes from conversational dynamics within the first 5 minutes. *J. Applied Psychology* 92, 3 (2007).

[23] Timothy DeGroot and Janaki Gooty. 2009. Can nonverbal cues be used to make meaningful personality attributions in employment interviews? *J. Business and Psychology* 24, 2 (2009).

[24] Timothy DeGroot and Stephan J Motowidlo. 1999. Why visual and vocal interview cues can affect interviewers' judgments and predict job performance. *J. Applied Psychology* 84, 6 (1999), 986.

[25] Christian M End and Katherine Saunders. 2013. Short Communication: Powerless and Jobless? Comparing the Effects of Powerless Speech and Speech Disorders on an Applicant's Employability. *Frontiers* 2, 1 (2013).

[26] Ray J Forbes and Paul R Jackson. 1980. nonverbal behaviour and the outcome of selection interviews. *J. Occupational Psychology* 53, 1 (1980), 65–72.

[27] David C Funder. 2006. Towards a resolution of the personality triad: Persons, situations, and behaviors. *J. Research in Personality* 40, 1 (2006), 21–34.

[28] Michelle Fung, Yina Jin, RuJie Zhao, and Mohammed Ehsan Hoque. 2015. ROC speak: semi-automated personalized feedback on nonverbal behavior from recorded videos. In *Proc. ACM UBIComp*. ACM, 1167–1178.

[29] Mark Gabbott and Gillian Hogg. 2000. An empirical investigation of the impact of non-verbal communication on service evaluation. *European J. Marketing* 34, 3/4 (2000), 384–398.

[30] Venkata Rama Kiran Garimella, Abdulrahman Alfayad, and Ingmar Weber. 2016. Social media image analysis for public health. In *Proc. 2016 CHI Conf. on Human Factors in Computing Systems*. ACM, 5543–5547.

[31] Daniel Gatica-Perez. 2009. Automatic nonverbal analysis of social interaction in small groups: A review. *Image and Vision Computing* 27, 12 (2009).

[32] Gerald J Gorn, Marvin E Goldberg, and Kunal Basu. 1993. Mood, awareness, and product evaluation. *J. Consumer Psychology* 2, 3 (1993), 237–256.

[33] Judith A Hall, Debra L Roter, and Cynthia S Rand. 1981. Communication of affect between patient and physician. *J. Health and Social Behavior* (1981), 18–30.

[34] Marvin A Hecht and Marianne LaFrance. 1995. How (Fast) Can I Help You? Tone of Voice and Telephone Operator Efficiency in Interactions1. *J. Applied Social Psychology* 25, 23 (1995), 2086–2098.

[35] James G Hollandsworth, Robert C Glazeski, and Mary Edith Dressel. 1978. Use of social-skills training in the treatment of extreme anxiety and deficient verbal skills in the job-interview setting. *J. Applied Behaviour Analysis* 11, 2 (1978).

[36] James G Hollandsworth, Richard Kazelskis, Joanne Stevens, and Mary Edith Dressel. 1979. Relative contributions of verbal, articulative, and nonverbal communication to employment decisions in the job interview setting. *J. Personnel Psychology* 32, 2 (1979).

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

[38] Jack L Howard and Gerald R Ferris. 1996. The employment interview context: Social and situational influences on interviewer decisions. *J. Applied Social Psychology* 26, 2 (1996), 112–136.
[39] Allen I Huffcutt, James M Conway, Philip L Roth, and Nancy J Stone. 2001. Identification and meta-analytic assessment of psychological constructs measured in employment interviews. *J. Applied Psychology* 86, 5 (2001).
[40] Hayley Hung, Yan Huang, Gerald Friedland, and Daniel Gatica-Perez. 2011. Estimating dominance in multi-party meetings using speaker diarization. *IEEE Transactions on Audio, Speech, and Language Processing* 19, 4 (2011), 847–860.
[41] 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* 62, 3 (1977).
[42] Hyo Sun Jung and Hye Hyun Yoon. 2011. The effects of nonverbal communication of employees in the family restaurant upon customer's emotional responses and customer satisfaction. *Int. J. Hospitality Management* 30, 3 (2011), 542–550.
[43] Juhee Kang and Sunghyup Sean Hyun. 2012. Effective communication styles for the customer-oriented service employee: Inducing dedicational behaviors in luxury restaurant patrons. *Int. J. Hospitality Management* 31, 3 (2012), 772–785.
[44] Douglas T Kenrick and David C Funder. 1988. Profiting from controversy: Lessons from the person-situation debate. *American psychologist* 43, 1 (1988), 23.
[45] Mark Knapp, Judith Hall, and Terrence Horgan. 2013. *Nonverbal communication in human interaction.* Cengage Learning.
[46] Max Kuhn. 2016. A Short Introduction to the caret Package. (2016).
[47] Nathan R Kuncel, Sarah A Hezlett, and Deniz S Ones. 2004. Academic performance, career potential, creativity, and job performance: Can one construct predict them all?
[48] Quoc Le and Tomas Mikolov. 2014. Distributed representations of sentences and documents. In *Proc. 31st ICML*. 1188–1196.
[49] Bruno Lepri, Nadia Mana, Alessandro Cappelletti, and Fabio Pianesi. 2009. Automatic prediction of individual performance from thin slices of social behavior. In *Proc. 17th ACM MMProc.* ACM, 733–736.
[50] Haim Mano and Richard L Oliver. 1993. Assessing the dimensionality and structure of the consumption experience: evaluation, feeling, and satisfaction. *J. Consumer research* 20, 3 (1993), 451–466.
[51] Thomas V McGovern and Howard EA Tinsley. 1978. Interviewer evaluations of interviewee nonverbal behavior. *J. Vocational Behavior* 13, 2 (1978).
[52] Matthias R Mehl, Megan L Robbins, and Fenne große Deters. 2012. Naturalistic observation of health-relevant social processes: the electronically activated recorder (EAR) methodology in psychosomatics. *Psychosomatic Medicine* 74, 4 (2012), 410.
[53] Stephan J Motowidlo and Jennifer R Burnett. 1995. Aural and visual sources of validity in structured employment interviews. *Organizational Behavior and Human Decision Processes* 61, 3 (1995), 239–249.
[54] Skanda Muralidhar and Daniel Gatica-Perez. 2017. Examining Linguistic Content and Skill Impression Structure for Job Interview Analytics in Hospitality. In *Proc. 16th ACM MUM*.
[55] 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*. 84–91.
[56] 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*.
[57] 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).
[58] 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* 16, 4 (2014).
[59] NIST NIST. 2012. Machine translation evaluation official results.
[60] Christopher Y Olivola, Dawn L Eubanks, and Jeffrey B Lovelace. 2014. The many (distinctive) faces of leadership: Inferring leadership domain from facial appearance. *The Leadership Quarterly* 25, 5 (2014), 817–834.
[61] Maja Pantic, Alex Pentland, Anton Nijholt, and Thomas S Huang. 2007. Human computing and machine understanding of human behavior: A survey. In *Artifical Intelligence for Human Computing*. Springer, 47–71.
[62] Charles K Parsons and Robert C Liden. 1984. Interviewer perceptions of applicant qualifications: A multivariate field study of demographic characteristics and nonverbal cues. *J. Applied Psychology* 69, 4 (1984), 557.
[63] Karl Pearson. 1901. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and J. Science* 2, 11 (1901), 559–572.
[64] James W Pennebaker and Laura A King. 1999. Linguistic styles: language use as an individual difference. *J. Personality and Social Psychology* 77, 6 (1999), 1296.
[65] Alex Pentland. 2004. Social dynamics: Signals and behavior. In *Int. Conf. on Developmental Learning*, Vol. 5.
[66] Alex Pentland and Tracy Heibeck. 2010. *Honest signals: how they shape our world.* MIT press.
[67] Bogdan Raducanu, Jordi Vitria, and Daniel Gatica-Perez. 2009. You are fired! nonverbal role analysis in competitive meetings. In *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE Int. Conf. on.* IEEE, 1949–1952.

[68] Keith G Rasmussen. 1984. Nonverbal behavior, verbal behavior, resumé credentials, and selection interview outcomes. *J. Applied Psychology* 69, 4 (1984), 551.
[69] Radim Řehůřek and Petr Sojka. [n. d.]. Software Framework for Topic Modelling with Large Corpora. In *Proc. LREC 2010 Workshop on New Challenges for NLP Frameworks*. ELRA.
[70] Dairazalia Sanchez-Cortes, Oya Aran, Marianne Schmid Mast, and Daniel Gatica-Perez. 2012. A nonverbal behavior approach to identify emergent leaders in small groups. *Multimedia, IEEE Trans. on* 14, 3 (2012), 816–832.
[71] Dairazalia Sanchez-Cortes, Petr Motlicek, and Daniel Gatica-Perez. 2012. Assessing the impact of language style on emergent leadership perception from ubiquitous audio. In *Proc. 11th Int. Conf. on MUM*. ACM, 33.
[72] Klaus R Scherer. 1982. Methods of research on vocal communication: Paradigms and parameters. *Handbook of Methods in Nonverbal Behavior Research* (1982), 136–198.
[73] Albrecht Schmidt. 2016. Cloud-based AI for pervasive applications. *IEEE Pervasive Computing* 15, 1 (2016), 14–18.
[74] Patrick E Shrout and Joseph L Fleiss. 1979. Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin* 86, 2 (1979).
[75] Aron W Siegman. 1987. The telltale voice: Nonverbal messages of verbal communication. (1987).
[76] Vivek K Singh, Souvick Ghosh, and Christin Jose. 2017. Toward Multimodal Cyberbullying Detection. In *Proc. 2017 CHI Conf. Extended Abstracts on Human Factors in Computing Systems*. ACM, 2090–2099.
[77] Del A Sole. 2018. *Getting Started with the Computer Vision API*. Apress, Berkeley, CA.
[78] Del A Sole. 2018. *Introducing Microsoft Cognitive Services*. Apress, Berkeley, CA.
[79] Brian R Spisak, Allen E Grabo, Richard D Arvey, and Mark van Vugt. 2014. The age of exploration and exploitation: Younger-looking leaders endorsed for change and older-looking leaders endorsed for stability. *The Leadership Quarterly* 25, 5 (2014), 805–816.
[80] DS Sundaram and Cynthia Webster. 2000. The role of nonverbal communication in service encounters. *J. Services Marketing* 14, 5 (2000), 378–391.
[81] Yla R Tausczik and James W Pennebaker. 2010. The psychological meaning of words: LIWC and computerized text analysis methods. *J. Language and Social Psychology* 29, 1 (2010), 24–54.
[82] Cornelis J Van Rijsbergen. 1986. A new theoretical framework for information retrieval. In *ACM SIGIR Forum*, Vol. 21. ACM.
[83] Chockalingam Viswesvaran and Deniz S Ones. 2000. Perspectives on models of job performance. *Int. J. Selection and Assessment* 8, 4 (2000), 216–226.
[84] Mark Weiser. 2002. The computer for the 21st century. *IEEE Pervasive Computing* 1, 1 (2002), 19–25.

Received February 2018; revised May 2018; accepted September 2018