# How May I Help You? Behavior and Impressions in Hospitality Service Encounters

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

Marianne Schmid Mast
University of Lausanne, Switzerland
Marianne.SchmidMast@unil.ch

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

## ABSTRACT

In the service industry, customers often assess quality of service based on the behavior, perceived personality, and other attributes of the front line service employees they interact with. Interpersonal communication during these interactions is key to determine customer satisfaction and perceived service quality. We present a computational framework to automatically infer perceived performance and skill variables of employees interacting with customers in a hotel reception desk setting using nonverbal behavior, studying a dataset of 169 dyadic interactions involving students from a hospitality management school. We also study the connections between impressions of Big-5 personality traits, attractiveness, and performance of receptionists. In regression tasks, our automatic framework achieves $R^2 = 0.30$ for performance impressions using audio-visual nonverbal cues, compared to 0.35 using personality impressions, while attractiveness impressions had low predictive power. We also study the integration of nonverbal behavior and Big-5 personality impressions towards increasing regression performance ($R^2 = 0.37$).

## CCS CONCEPTS

• **Applied computing → Psychology**; • **Human-centered computing → *Empirical studies in HCI***; Empirical studies in ubiquitous and mobile computing;

## KEYWORDS

Social computing; first impression; hospitality; nonverbal behavior; multimodal interaction; job performance; reception desk

**ACM Reference Format:**
Skanda Muralidhar, Marianne Schmid Mast, and Daniel Gatica-Perez. 2017. How May I Help You? Behavior and Impressions in Hospitality Service Encounters. In *Proceedings of 19th ACM International Conference on Multimodal Interaction (ICMI'17).* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3136755.3136771


*ICMI'17, November 13–17, 2017, Glasgow, UK*

ACM ISBN 978-1-4503-5543-8/17/11…$15.00
https://doi.org/10.1145/3136755.3136771

## 1 INTRODUCTION

The interaction between service employees (e.g. reception desk assistants) and customers, commonly referred to as service encounters in the hospitality industry, is a critical part of customer experience at an establishment [51]. It is during these encounters that customers perceive and evaluate the employee's attitudes and professional, social, and communication skills. Based on these interactions, customers form impressions of both the employee and the organization [37]. The importance of interpersonal communication during service encounters in determining customer satisfaction and perceived quality of service (QoS) has been highlighted in prior work [18, 51]. Literature in psychology, marketing, and hospitality has demonstrated that as a major component of interpersonal communication [31], nonverbal behavior (NVB) contributes towards shaping the outcome of customer-employee interactions [17, 51]. Customers often use interactions with front line service employees to assess QoS [24], it is imperative for hospitality organizations to improve the quality of these encounters. In this work, bringing audio-visual processing and machine learning as additional analytical tools, we investigate the connections between automatically extracted nonverbal behavior and performance impressions in service encounters, in addition to other important factors including employees' personality traits and attractiveness. Specifically, we study dyadic interactions at a hotel reception desk setting between employees and customers.

Job performance is a central construct in organizational psychology and has a variety of definitions in the literature [52]. Specific aspects of the performance construct may change from job to job, but some dimensions can be generalized across jobs (e.g. interpersonal communication). In this work, we use the definition proposed in [52], which denotes job performance as "action, behavior and outcomes that employees engage in and contribute to organizational goals". We define *performance impressions* as the behavioral aspect of performance as perceived by others who can observe an interaction (e.g. in a dyadic service encounter) and assess the performance of the employee based on the interaction itself.

In this paper, we study the connections between performance and related impressions and automatically extracted nonverbal behavior, as well as other relevant variables discussed in hospitality, marketing, and psychology literature, namely perceived personality traits and attractiveness. Our work builds upon and extends work on automatic analysis of social interaction in the workplace [19, 42], which has shown the potential of inferring negotiation outcomes [15], job interview ratings [13, 40], and other constructs (e.g. engagement, friendliness, or excitement) [25, 39, 47] up to a certain level of performance. We study a dataset consisting of 169 dyadic interactions between a hotel desk receptionist and a client, where receptionists are played by students from an international hospitality management school who practice real-life situations. We address the following research questions:

**RQ1:** What nonverbal cues displayed by desk service employees are connected to performance and skill impressions? Can they be used to automatically infer these constructs?

**RQ2:** How are perceived Big-5 personality traits of desk service employees linked to performance impressions?

**RQ3:** Are there any connections between the perceived attractiveness of such employees and their performance impressions?

The contributions of this work are the following. First, we analyze the relationship between automatically extracted audio-visual nonverbal cues and performance and skill impressions via correlation analysis and a regression task. Interestingly, we show that the customer's nonverbal cues explain up to 27% of the variance of the participant's perceived performance scores. Second, we show that Big-5 trait impressions achieve performance of $R^2 = 0.35$ for job performance impressions. Third, we show that judgments of attractiveness were not good predictors of impression and skill ratings. Finally, the integration of NVB cues and Big-5 impression results in an overall benefit in inference of performance impressions. This research might have wider implications for employees and managers in hospitality, by providing an understanding of not only the employee's performance via nonverbal behavior, but also of the implications for customer perceptions of service encounters. The automatic approach could also facilitate personalized training for employees to improve their nonverbal behavior in service encounters.

# 2 PREVIOUS WORK

## 2.1 Literature in Psychology

First impressions are defined as formation of a mental image of a person when met for the first time [5]. Research in psychology has revealed that nonverbal behavior is an important component in the formation of first impressions [31, 48] and that even a short interaction ("Thin slices") is enough to form first impressions [3]. Specifically in the workplace setting, thin slices of nonverbal behavior have shown to be predictive of job interviews [41], evaluation of sales job performance [2], and employee-customer interaction [6]. Regarding assessment of performance, Ambady et al. showed that end of semester ratings of 13 university teachers as rated by students could be predicted based on judgments of personality characteristics from 10-second clips [4]. The same authors showed the predictive validity of thin slice judgments on the performance of 12 sales managers using 30-sec audio clips [2].

The influence of physical attractiveness on impressions has been investigated in psychology. Ahearne et al. reported a positive effect of physical attractiveness on sales performance of 339 pharmaceutical sales representatives [1]. Here, attractiveness was rated by physicians based on their personal interaction with sales representatives. Similarly, Hamermesh et al. reported that college professors perceived as physically attractive were evaluated higher by students [23]. In this study, six undergraduate students (3 females) rated perceived attractiveness of 94 professors using photos posted on department websites. Magnus et al. investigated the physical attractiveness of service workers in two settings (bookstores and airline travel) and reported the positive impact of physical attractiveness on customer satisfaction in both settings [50]. In this study, attractiveness was rated based on photos of the service workers.

## 2.2 Literature in Hospitality and Marketing

In the context of hospitality, sales and marketing, and management, the relationship between impression formation and nonverbal behavior has also been acknowledged [11, 51]. Many aspects of nonverbal behavior including gestures, smiles, touch, and prosody as well as other attributes like physical attractiveness have been explored. Gabbott and Hogg, using a study conducted using a video recording of an actress playing the role of a reception desk assistant helping a customer to check-in, showed that nonverbal communication impacts the customer perception of QoS [17]. Furthermore, the study also showed the effect of perceived attractiveness on satisfaction of service. Kang and Hyun investigated the effect of communication styles on customer-oriented service in a study consisting of 527 luxury restaurant patrons [30]. In this study, the authors emailed a questionnaire to participants based on their visit to a luxury restaurant. They reported that employees who smiled, nodded and maintained eye contact with the clients, and spoke with high energy and tone of voice with fewer short utterances were positively correlated to customer satisfaction.

Jung and Yoon, in a study consisted of 333 customers, investigated the role of nonverbal behavior in customer satisfaction at a family restaurant in South Korea [29]. The authors reported 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 ($r = 0.33; p < 0.01$).

The relationship between personality and job performance impressions was investigated in [7]. The authors conducted a meta-analysis and reported correlations between Big-5 personality traits and job performance. Specifically, they found that *Conscientiousness* was correlated to all occupational groups in the study, while *Extraversion* was found to be a valid predictor for jobs that require social interactions, like managers.

The effect of physical attractiveness has been a subject of interest in the marketing and service industry. In hospitality, it was shown that tips received by female waitresses from male customers were positively related to service providers wearing makeup [27] and certain colored clothes [22]. Both these studies were conducted in the field, and attractiveness was rated based on physical appearance during the interaction in a restaurant setting. Similarly, Luoh et al. reported that customer's perceptions of service quality were enhanced by attractive service providers compared to those of average appearance [34]. A "Beauty is beastly" effect was reported in a study which found that physical attractiveness for women was detrimental in employment contexts considered to be masculine (e.g managers, director of security etc) [28]. In both studies [28, 34], attractiveness was rated based on photographs controlled for background, age, and posture.

Most research in social psychology, hospitality, and marketing relies on manual annotations of nonverbal behavior, making it labor-intensive and difficult to scale with respect to large number of users and different scenarios. The advent of ubiquitous sensors combined with improved perceptual techniques have enabled the automatic analysis of human interactions [19, 42].

## 2.3 Literature in Computing

Literature in social computing has validated the viability to integrate nonverbal behavior extracted using ubiquitous sensors and machine learning algorithms to infer various constructs like interview ratings [39, 40], negotiation outcomes [15], and Big-5 personality [8] to promising levels [20].

Batrinca et al. used an approach to predict Big-5 traits from self-presentation questions, where participants introduced themselves in front of a computer, similar to job interviews, but without the presence of an interviewer [8]. Nguyen et al. used automatically extracted nonverbal cues (speaking turns, prosody, head nods, visual activity) from both applicant and interviewer to infer five hirability variables in a dataset consisting of 62 real job interviews [40]. Naim et al. extended these works by analyzing a dataset of 138 simulated job interviews from internship-seeking students [39]. The authors extracted cues related to facial expressions, verbal content, and prosody to predict several variables (hiring recommendation, engagement, friendliness). Chen et al. developed a standardized video interview protocol along with human ratings, which focused on verbal content, personality, and holistic judgment [13]. The authors using "visual words" as feature extraction method, automatically learned from video analysis outputs, and the Doc2Vec paradigm achieved a correlation of 0.42 between machine-predicted scores and human-rated scores.

Biel et al. studied effects of physical attractiveness in a study focusing on modeling different facets of YouTube vloggers [10]. Using 442 vlogs rated for two dimensions of physical attractiveness, and three dimensions of non-physical attractiveness, they reported significant positive correlations between judgments of attractiveness and two Big-5 traits (Extraversion and Agreeableness). Attractiveness was rated on 1-min vlogs by Amazon Mechanical Turk workers.

In the context of performance, Curhan et al. investigated the relationship between audio cues and negotiation outcomes in dyadic job negotiations [15]. Performance was measured as the compensation package that could be negotiated. The authors reported that voice activity levels, prosodic emphasis, and vocal mirroring explained up to 30% of the variance. Madan et al reported the validity of audio nonverbal cues in predicting the performance of male participants in a speed dating setup [35]. In a study consisting of 57 five-minute sessions, the authors reported positive correlation between a female 'liking' a male participant and aggregated male and female speech features ($r = 0.67; p < 0.05$). In this setup, performance was measured as the number of likes received from female participants. Lepri et al. investigated the use of nonverbal behavior for inference of individual performance in a group task [33]. Using audio and visual features extracted from the Mission Survival 2 corpus, they were able to classify binary levels of performance with accuracy up to 50%. Raducanu et al. used a dataset collected from the reality show "The Apprentice" to predict the person who will be fired [44]. Using speaking turn features, the method predicted the candidate to be fired ( i.e the one with worst perceived performance) with an accuracy of 92%.

The existing literature in psychology and social computing demonstrates the feasibility of predicting up to a certain degree either actual performance or performance impressions using thin slices of nonverbal behavior and other constructs like attractiveness and personality. In this work, we investigate the interplay between nonverbal behavioral cues, and impressions of attractiveness, personality traits, and performance in a novel setting. To the best of our knowledge, there has been no studies to automatically infer performance between reception desk assistants and customers in a hospitality setting. Our work, therefore, could have wider implications for managers and training students in the fields of customer service and hospitality.

**Figure 1: The reception desk setting including (1) the client (research assistant, on the right), (2) the receptionist (participant, on the left), (3) microphone array, and (4) two Kinect devices.**

## 3 RECEPTION DESK DATASET

The reception desk is considered the entry point of a hotel and, as a very frequent type of interaction in hospitality, is often determinant of the evaluations of service quality of such organizations [24, 46]. Despite its importance, there is no publicly available dataset to study this interaction from the perspective of performance impressions. We use a dataset of 169 dyadic interactions in a reception desk setting [38]. This dataset was described but not used in our previous work [38] and is studied for the first time here.

### 3.1 Scenario and Data Collection

The reception desk dataset was collected as part of a multi-situation corpus involving hospitality school students taking part in job interviews and reception desk in two practice sessions, for either two (one interview, one desk) or four (two interviews, two desk) interactions.

The reception desk role-play involves a receptionist (a hospitality student participant) and a client (a research assistant selected from a seven-person group of master's students in business and psychology). The two protagonists discuss at the reception desk of a hotel. A snapshot is presented in Figure 1. The scenario is the following: *The participant is an intern at a high-end hotel and is assigned to the reception desk. As the reception manager is not available, the participant has to handle all customer interactions. A client comes to check out. The initial attitude of the client is friendly. The bill is then handed to the client. Once the customer reads the bill, his/her attitude changes as (s)he complains about bill-related problems (costs associated to taxes, TV, and WiFi). The participant needs to handle the* *situation and come to a resolution that is acceptable for the customer.* The goal of the scenario is to elicit a situation that allows to assess how the participants playing the receptionist role handle real-life interactions, in order to assess their performance at the situation. The receptionist does not know what the exact client's reaction will be. For students participating in a second session, a variation of the scenario was used to reduce predictability, in which a new client (played by a different research assistant) changed her/his attitude before receiving the bill, complaining about a bad restaurant recommendation by a previous receptionist.

Video data was collected with two Kinect v2 devices (one for client, one for receptionist), each recording 30 fps RGB+depth video ($1920 \times 1080$ and $512 \times 424$ for RGB and depth, respectively.) Audio data was collected using a microphone array device, that captures audio at 48kHz and segments speaker turns from localized sources. The audio and video streams are synchronized in a subsequent step.

Study participants were students at an international hospitality management school. A total of 100 students voluntarily took part in the study (mean age 20.6 years old; 57 females and 43 males). 69 participants contributed two reception desk interactions. Interactions were allowed in either English or French (to the choice of each participant) to reflect the international population of the hospitality school, resulting in 130 (resp. 39) interactions in French (resp. English). As a whole, the Reception Desk dataset has a total duration of 1350 minutes (mean duration: 8 mins).

## 3.2 Annotations

We enriched the audio-visual dataset with a number of manually labeled variables. Impressions of performance, skills and personality traits (Big-5) were coded by one group consisting of three independent annotators (Group-A), while attractiveness attributes were rated by a separate group of three independent annotators (Group-B). The choice of two separate groups was motivated by the fact that asking the raters to focus on physical attractiveness could influence performance impressions, the very question under study. The annotators were students, who responded to a call for volunteers and were paid 20 USD per hour for their work.

Annotators in Group-A watched the first two minutes of the complaint segment of all the reception desk videos. These two-minute segments were selected as thin slices, following previous work in psychology [3] and social computing [33, 43]. Annotators rated all the receptionists on a number of impression variables using a seven-point Likert scale. The list of variables (Table 1) includes: *Performance* (participant's ability to stay calm, satisfy customers, be patient and calm, and be resistant to stress); *Overall Impression*; *Professional Skills* (Competent, Motivated, Satisfying); *Social Skills* (Intelligent, Positive, Sociable); and *Communication Skills* (Clear, Persuasive). Several of these variables have been studied in other work-related computational studies [38–40]; we examine them for the reception desk case. In addition, the annotators were asked to rate the perceived personality traits of all participants. We used the standard Ten Item Personality Inventory (TIPI) consisting of ten items, two per dimension [21]. TIPI is widely used to collect impressions of personality and is easy to administer as it consists of only ten questions. Our goal is not to predict personality, but to use big-5 traits as features to predict job performance, following previous literature.

**Table 1: Reception desk dataset: $ICC(2, k)$ & descriptive statistics for impressions of skills and performance, attractiveness & personality traits.**

| Variable | ICC | mean | std | median | skew |
|---|---|---|---|---|---|
| *Professional Skills* | | | | | |
| Competent (compe) | 0.69 | 4.24 | 1.36 | 4.33 | -0.30 |
| Motivated (motiv) | 0.63 | 4.80 | 1.12 | 5.00 | -0.46 |
| Satisfying (satis) | 0.73 | 4.16 | 1.41 | 4.33 | -0.15 |
| *Social Skills* | | | | | |
| Intelligent (intel) | 0.58 | 4.52 | 1.04 | 4.67 | -0.18 |
| Positive (posit) | 0.60 | 4.34 | 1.09 | 4.33 | -0.07 |
| Sociable (socia) | 0.64 | 4.46 | 1.14 | 4.33 | -0.26 |
| *Communication Skills* | | | | | |
| Clear (clear) | 0.66 | 4.56 | 1.25 | 4.67 | -0.53 |
| Persuasive (persu) | 0.72 | 4.01 | 1.38 | 4.00 | -0.07 |
| *Overall* | | | | | |
| Overall Impression (ovImp) | 0.75 | 4.27 | 1.46 | 4.33 | -0.13 |
| Performance (peImp) | 0.77 | 4.11 | 1.37 | 4.33 | -0.06 |
| *Big-5 Personality* | | | | | |
| Agreeableness (agree) | 0.62 | 3.5 | 0.59 | 3.5 | 0.14 |
| Conscientiousness (consc) | 0.41 | 3.9 | 0.30 | 4.0 | -0.44 |
| Extraversion (extra) | 0.68 | 4.23 | 0.41 | 4.33 | -0.14 |
| Neuroticism (neuro) | 0.47 | 4.11 | 0.38 | 4.0 | 0.35 |
| Openness (open) | 0.40 | 4.19 | 0.29 | 4.17 | 0.71 |
| *Attractiveness* | | | | | |
| Attractiveness (attrac) | 0.62 | 3.72 | 1.44 | 3.73 | 0.27 |
| Dislikeable (disli) | 0.36 | 3.96 | 1.26 | 4.01 | -0.18 |
| Friendly (frien) | 0.59 | 3.77 | 1.43 | 3.41 | 0.14 |
| Likeable (likea) | 0.55 | 3.48 | 1.35 | 3.41 | 0.21 |

Annotators in Group-B were asked to rate attractiveness of the participants based on still images. The images were video frames selected based on the following criteria: (1) Full frontal face was visible with no occlusion, and (2) the participant displayed a neutral face ( no smiling or any other expression). The attractiveness of each participant was assessed using four variables: *Physical Attractiveness*, *Likeable*, *Dislikeable*, and *Friendly*. The use of both physical and non-physical attractiveness was inspired by its previous use in literature [1, 28, 34, 45, 50]. Annotators were asked to answer the questions: "How attractive do you find this person?" for *Physical Attractiveness*, and similarly for the other attributes. This was rated on a five-point Likert scale which was later rescaled to seven-point scale before analysis. The ratings of six participants were excluded due to technical reasons thus rendering $N = 163$ for all attractiveness related analysis.

Table 1 summarizes the descriptive statistics and the agreement between raters measured using the Intraclass Correlation Coefficient (ICC), a commonly used metric in psychology [49]. We observe that the distribution of all performance and skill variables and three personality traits are centered on the positive side of the Likert scales (Mean $\geq$ 4 ). In contrast, the other personality traits (*Agreeableness* and *Conscientiousness*) and attractiveness attributes are on the negative side of the Likert scale (Mean $\leq$ 4 ). We used $ICC(2, k)$ as measure of inter-rater agreement, given that (1) a sample (rather than a population) of annotators was used, and (2) each annotator judged all videos (images in the case of attractiveness). Agreement between raters for performance and skill impression variables was moderate to high, with $ICC(2, k) \in [0.58, 0.77]$.

**Table 2: Nonverbal features extracted from the reception desk data.**

| Features | | | |
|---|---|---|---|
| Speech Activity | Prosody | Visual | Multimodal |
| Speaking time | Pitch | Overall visual motion | Speaking while nodding |
| Speaking turns | Speaking Rate | Head nods | |
| Pauses | Spectral Energy | Visual Back-channelling | |
| Short Utterance | Speaking Energy | | |
| Silence | Voicing Rate | | |
| | Rate of change of energy | | |

Similarly, the agreement between raters for personality traits impression $[0.41 < ICC(2,k) < 0.68]$ and attractiveness attributes $[0.36 < ICC(2,k) < 0.62]$ was moderate. For Big-5, this could be due to the interaction setting, which elicits *Agreeableness* and *Extraversion* traits to be more visible. Overall, the inter-rater agreement for impressions of personality traits and attractiveness are in similar range to those reported in literature [10].

# 4 NONVERBAL FEATURES

To understand the influence of nonverbal behavior on the formation of performance and skill impressions, various cues were automatically extracted from the audio and visual modalities from both the receptionist and client. The choice of nonverbal cues were guided by their relevance in existing literature in social psychology [2, 16, 26] and social computing [33, 38, 40]. The nonverbal cues were extracted from the moment the client gets the bill and changes to an unfriendly attitude until the end of the interaction.

**Speaking Activity Features** include speaking time (total time an individual speaks), speaking turns (segments greater than two seconds), pauses (gaps in speech shorter than two seconds), short utterances (speaking segments less than two seconds). The cues were extracted based on the speaker diarization provided by the commercial microphone array. These cues are known to have a connection to impression formation in various workplace interactions [33, 40]. Various statistics like count, mean, standard deviation, maximum, and minimum values were calculated and used as features.

**Prosody 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). They were extracted using existing code [36]. The following statistics were extracted and used as features: mean, standard deviation, minimum, maximum, entropy, median, and quartiles.

**Head Nods** are periodic up-and-down head movements. A 3D face centered method was used towards this objective [14]. In this method, a 3D head tracker calculates the angular velocities using relative rotation at each instant with respect to the head pose at some earlier instance. Count, mean, median, standard deviation, minimum, and maximum of head nods duration were computed as features.

**Visual Back-Channeling** (visual BC) are events when a person nodded while the other was speaking. Number of nods, mean, median, standard deviation, minimum, and maximum of visual BC duration were computed as features.

**Overall Visual Motion** was computed by a modified version of motion energy images, called Weighted Motion Energy Images (WMEI) [9]. This captures the total amount of visual movement displayed by both receptionist and client. Statistics of WMEI include mean, median, standard deviation, minimum, maximum, entropy, quartiles, and center of gravity.

**Table 3: Correlation matrix for perceived performance & skill impressions in reception desk ($N = 169$) with significance value ($p < 0.001$) in all cases.**

| Impressions | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|
| 1.peImp | .96 | .94 | .82 | .92 | .90 | .82 | .78 | .89 | .92 |
| 2.ovImp | | .92 | .84 | .91 | .88 | .85 | .81 | .87 | .90 |
| 3.compe | | | .80 | .93 | .92 | .82 | .77 | .89 | .92 |
| 4.motiv | | | | .77 | .77 | .74 | .79 | .71 | .77 |
| 5.satis | | | | | .89 | .81 | .72 | .87 | .94 |
| 6.intel | | | | | | .79 | .75 | .87 | .89 |
| 7.posit | | | | | | | .85 | .77 | .80 |
| 8.socia | | | | | | | | .71 | .73 |
| 9.clear | | | | | | | | | .91 |
| 10.persu | | | | | | | | | |

**Multimodal Cues** are defined as events when protagonists nod their head while speaking. Count of nodding while speaking, mean, median, standard deviation, minimum, and maximum of duration were computed for use as features.

All the features extracted are summarized in Table 2 and were normalized with respect to the interaction duration. As part of future work, we plan to study additional features including smiling, gaze, and verbal content.

# 5 CORRELATION ANALYSIS

This section presents the Pearson correlation analysis performed to understand performance and skill impressions in this setup and their relationship with nonverbal cues, personality trait impressions, and attractiveness. For the analysis, the average of each impression variable provided by the three raters is used.

As a first step, correlation between the skill and performance variables was computed (Table 3). All variables annotated are significantly correlated with each other, with correlation coefficients ($r$) greater than .60 for most cases. Correlations between some variables are very high like *performance impression* and *overall impression* ($r = .96$), indicating that they are essentially measuring the same construct. For all subsequent analyzes, *overall impression* (ovImp) is not included due to this fact.

## 5.1 Performance Impressions and Nonverbal Cues (RQ1)

In the next step, correlations between extracted nonverbal cues and annotated variables were investigated. Correlation of selected nonverbal cues are presented in Table 4. A number of receptionist's features were found to be significantly correlated to impressions of performance and skills. Specifically, receptionists who spoke for longer duration, faster, took longer turns, and had fewer silence events obtained higher scores for performance and skill impressions. Similarly, receptionists who spoke animatedly with higher visual motion, nodded more, displayed greater visual BC, were more favorably viewed than those who spoke with less visual activity.

Literature in psychology and hospitality has reported that the use of faster speech, with fewer silent events, enhances customer's perception of competence, while more head nods and visual BC enhances the perception of empathy, courtesy, and trust [51]. Our

**Table 4: Pearson correlation between nonverbal cues and performance & skill impressions.** $N = 169$; $^{*}p < 0.001$, $^{\dagger}p < 0.01$, $^{**}p < 0.05$.

| NVB Cues | Skills | | | peImp |
|---|---|---|---|---|
| | Professional | Social | Communication | |
| *Receptionist* | | | | |
| *Speaking Activity* | | | | |
| Speaking Ratio | $[.35, .43]^{\dagger}$ | $[.37, .44]^{\dagger}$ | $[.30, .39]^{\dagger}$ | $.43^{\dagger}$ |
| Mean Turn Duration | $[.32, .37]^{\dagger}$ | $[.33, .38]^{\dagger}$ | $[.30, .34]^{\dagger}$ | $.40^{\dagger}$ |
| Max Turn Duration | $[.39, .41]^{\dagger}$ | $[.36, .39]^{\dagger}$ | $[.34, .41]^{\dagger}$ | $.42^{\dagger}$ |
| Num Silence Events | $[-.23, -.22]^{\dagger}$ | $[-.29, -.18]^{\dagger}$ | $[-.21, -.18]^{\dagger}$ | $-.22^{\dagger}$ |
| *Voicing Rate* | | | | |
| Mean | $[.31, .34]^{\dagger}$ | $[.32, .34]^{\dagger}$ | $[.32, .32]^{\dagger}$ | $.28^{\dagger}$ |
| Voicing Rate Q25 | $[.29, .32]^{\dagger}$ | $[.28, .35]^{\dagger}$ | $[.28, .29]^{\dagger}$ | $.28^{\dagger}$ |
| Voicing Rate Q75 | $[.27, .30]^{\dagger}$ | $[.27, .35]^{\dagger}$ | $[.28, .31]^{\dagger}$ | $.24^{\dagger}$ |
| *Visual Motion* | | | | |
| Mean WMEI | $[.19, .33]^{\dagger}$ | $[.18, .36]^{\dagger}$ | $[.20, .22]^{\dagger}$ | $.26^{\dagger}$ |
| Max WMEI | $[.30, .33]^{\dagger}$ | $[.26, .37]^{\dagger}$ | $[.31, .31]^{\dagger}$ | $.30^{\dagger}$ |
| Count Head Nods | $[.35, .42]^{\dagger}$ | $[.39, .41]^{\dagger}$ | $[.34, .41]^{\dagger}$ | $.37^{\dagger}$ |
| *Visual BC* | | | | |
| Count | $[.20, .29]^{\dagger}$ | $[.26, .27]^{\dagger}$ | $[.22, .25]^{\dagger}$ | $.23^{\dagger}$ |
| Mean Duration | $[.22, .22]^{\dagger}$ | $[.20, .25]^{\dagger}$ | $[.20, .20]^{**}$ | $.18^{**}$ |
| Max Duration | $[.25, .30]^{\dagger}$ | $[.26, .29]^{\dagger}$ | $[.23, .25]^{\dagger}$ | $.25^{\dagger}$ |
| Min Duration | $[-.26, -.19]^{\dagger}$ | $[-.26, -.18]^{\dagger}$ | $[-.24, -.19]^{**}$ | $-.22^{\dagger}$ |
| *Multimodal Cues* | | | | |
| Count | $[.44, .49]^{\dagger}$ | $[.45, .49]^{\dagger}$ | $[.41, .49]^{\dagger}$ | $.45^{\dagger}$ |
| Mean Duration | $[.23, .30]^{\dagger}$ | $[.22, .27]^{\dagger}$ | $[.26, .30]^{\dagger}$ | $.24^{\dagger}$ |
| Max Duration | $[.40, .43]^{\dagger}$ | $[.40, .43]^{\dagger}$ | $[.39, .44]^{\dagger}$ | $.39^{\dagger}$ |
| Min Duration | $[-.27, -.23]^{\dagger}$ | $[-.33, -.25]^{\dagger}$ | $[-.24, -.23]^{\dagger}$ | $-.24^{\dagger}$ |
| *Client* | | | | |
| *Voicing Rate* | | | | |
| Mean Voicing Rate | $[.24, .31]^{\dagger}$ | $[.30, .35]^{\dagger}$ | $[.23, .24]^{\dagger}$ | $.25^{\dagger}$ |
| Voicing Rate Q25 | $[.17, .23]^{\dagger}$ | $[.23, .26]^{\dagger}$ | $[.19, .21]^{\dagger}$ | $.19^{**}$ |
| Voicing Rate Q75 | $[.24, .27]^{\dagger}$ | $[.27, .29]^{\dagger}$ | $[.21, .22]^{\dagger}$ | $.21^{\dagger}$ |
| *Visual Motion* | | | | |
| Max WMEI | $[.30, .33]^{\dagger}$ | $[.26, .37]^{\dagger}$ | $[.31, .31]^{\dagger}$ | $.33^{\dagger}$ |
| Count Head Nod | $[.18, .30]^{\dagger}$ | $[.24, .29]^{\dagger}$ | $[.27, .31]^{\dagger}$ | $.24^{\dagger}$ |
| *Visual BC* | | | | |
| Count | $[.25, .30]^{\dagger}$ | $[.28, .30]^{\dagger}$ | $[.26, .31]^{\dagger}$ | $.30^{\dagger}$ |
| Max Duration | $[.20, .21]^{\dagger}$ | $[.22, .24]^{\dagger}$ | $[.17, .21]^{\dagger}$ | $.17^{**}$ |
| Min Duration | $[-.24, -.22]^{\dagger}$ | $[-.20, -.16]^{\dagger}$ | $[-.22, -.17]^{\dagger}$ | $-.21^{\dagger}$ |

results are comparable with previous literature for other conversational settings like interviews [16, 40], restaurant service [29], and sales [2].

An interesting insight is the correlation between some of the client's nonverbal cues and the impression score of the receptionist. We observed that clients tend to speak faster with greater visual motion in presence of receptionists who were rated higher. Also, clients tend to nod more and provide greater visual BC and for longer duration while interacting with receptionists who scored higher than with receptionists with lower scores. Similar results are reported in other dyadic settings like job interviews [38, 40].

## 5.2 Performance Impressions and Personality Trait Impressions (RQ2)

The correlation between Big-5 personality impressions and *performance impression* was computed (Table 5). *Extraversion* was observed to be positively correlated to *performance impression* with $r = 0.42$ ($p < 0.001$), while *Neuroticism* was found to be negatively correlated with $r = -0.39$ ($p < 0.001$). *Agreeableness* and *openness* were observed to have lower correlation to *performance impression* with $r = .23$ ($p < 0.01$) and $r = .15$ ($p < 0.05$) respectively. Interestingly, we do not observe any correlation between *conscientiousness* and *performance impression* as suggested in [7]. This could be explained as in this situation, *conscientiousness* is a hard trait to score and has lower agreement among raters ($ICC(2,k) = 0.41$). The results of other personality trait impressions are in line with literature in psychology, especially *Extraversion*, which is reported to be a valid predictor of performance for jobs requiring social interactions [7].

**Table 5: Correlation between Big-5 personality trait impressions and *Performance Impressions*** ($N = 169$; $^{*}p < 0.001$, $^{\dagger}p < 0.01$, $^{**}p < 0.05$)**. Entries without p-value symbol are not significant.**

| Impressions | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|
| 1.peImp | $.23^{\dagger}$ | $-.11$ | $.42^{*}$ | $-.39^{*}$ | $.15^{**}$ |
| 2.agree | | $-.05$ | $.08$ | $-.45^{*}$ | $-.01$ |
| 3.cons | | | $-.05$ | $.05$ | $-.03$ |
| 4.extra | | | | $-.02$ | $.08$ |
| 5.neuro | | | | | $-.08$ |
| 6.open | | | | | |

## 5.3 Performance Impressions and Attractiveness (RQ3)

The correlation analysis of attractiveness attributes and *performance impression* yielded unexpected results (Table 6). It was observed that *attractive* was not significantly correlated to *performance impression*, while *friendly* ($r = -0.27$) and *likeable* ($r = -0.26$) had low negatively correlation ($p < 0.05$). *Dislikeable* was observed to have low positive correlation ($r = 0.17; p < 0.05$). Given the literature on gender and attractiveness and job performance [28, 50], we divided the sample of receptionists based on gender. It was observed that for males ($N = 79$) there was no correlation between any attractiveness attributes and *performance impression* ($r \in [0.01, -0.05]$). However, for the female receptionists ($N = 90$), *friendly* ($r = -0.47$) and *likeable* ($r = -0.48$) was negatively correlated to *performance impression* ($p < 0.001$), while *dislikeable* was positively correlated ($r = 0.40; p < 0.001$). This result does not conform several of the results reported in the literature of attractiveness and performance, where a positive connection was often found [23, 50]. For further discussion, refer to Section 6.3.

# 6 REGRESSION ANALYSIS

A framework for inference of impressions of performance and skills from nonverbal cues, personality impressions, and attractiveness impressions was proposed and evaluated. The data was first preprocessed by a person-independent Z-score normalization to transform data into unity variance and zero mean. Then, both a full feature representation and standard dimensionality reduction techniques (Principle Component Analysis and significantly correlated features) were evaluated.

Regarding the machine learning approach, two regression techniques (Ridge Regression (Ridge) and Random Forest (RF)) implemented in the Caret R package were evaluated [32]. Leave-one-person-out cross-validation and 10-fold inner cross-validation were used. Hyper parameters (i.e., number of trees, shrinkage parameters) were automatically tuned by using an inner 10-fold cross-validation on the training set. Performance of these regression techniques were evaluated by employing two standard measures: coefficient of determination ($R^2$) and root-mean-square error ($RMSE$).

Here, results of only the best performing model are presented and discussed. For this task, as the baseline we use $R^2 = 0$ by predicting the population mean.

**Table 6: Correlation between *Attractiveness* attributes and *Performance Impression* ($N = 163$;$^* p < 0.001$,$^\dagger p < 0.01$, $^{**} p < 0.05$). Entries without p-value symbol are not statistically significant.**

| | All Receptionists | | | | Female Receptionists | | | | Male Receptionists | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Impressions | 2 | 3 | 4 | 5 | 2 | 3 | 4 | 5 | 2 | 3 | 4 | 5 |
| 1.peImp | $-.12$ | $.18^{**}$ | $-.27^{**}$ | $-.26^{**}$ | $-.18$ | $.40^*$ | $-.47^*$ | $-.48^*$ | $-.02$ | $-.05$ | $.08$ | $-.06$ |
| 2.attra | | $-.44^*$ | $.60^*$ | $.54^*$ | | $-.56^*$ | $.68^*$ | $.69^*$ | | $-.26^{**}$ | $.47^*$ | $.33^\dagger$ |
| 3.disli | | | $-.74^*$ | $-.78^*$ | | | $-.86^*$ | $-.85^*$ | | | $-.59^*$ | $-.72^*$ |
| 4.frien | | | | $.83^{**}$ | | | | $.89^*$ | | | | $.75^*$ |
| 5.likea | | | | | | | | | | | | |

## 6.1 Performance Impressions & Nonverbal Cues (RQ1)

Regression results indicate that all variables could be predicted to a certain degree from automatically extracted nonverbal cues (Table 7). It is observed that 30% of variance in *performance impression* (peImp) can be explained by nonverbal cues. Other variables have similar predictability using aggregated nonverbal cues of *Participant* and *Client*. Specifically, *sociable* (socia) has the highest performance ($R^2 = .33$), followed by *positive* (posit), *persuasive* (persu) (both $R^2 = .32$), and *motivated* (motiv) ($R^2 = .30$). These results provide an answer to RQ1: nonverbal behavior is predictive of performance and skill impressions in this hospitality encounter scenario. These results also corroborate findings in other conversational settings like job interviews [38, 40] and job negotiations [15]. In [15] the authors were able to explain up to 30% of the variance in job negotiation performance using audio features, while the authors in [38, 40] reported $R^2 = 0.34$ and $R^2 = 0.32$ for hirability and overall impressions respectively. Our results are in the same range. In hospitality literature, it has been shown that nonverbal behavior is correlated to customer satisfaction ($r \in [0.33, 0.42]$) [29]. We compare these results to this work by converting $r$ to $R^2$ (our evaluation measure, coefficient of determination $R^2$, is approximated by computing the square of correlation coefficient $r$). They reported a prediction accuracy of $r = 0.42$ for overall performance, which indicates a $R^2$ of 0.18. Some variables like *clear* (clear) ($R^2 = 0.22$) are harder to predict using extracted nonverbal cues. Our results can be seen as baseline for this type of task in the reception desk setting, and the reported $R^2$ is comparable with results obtained in other tasks in the literature.

As a next step, the contribution of the nonverbal cues from each protagonist was investigated. Receptionist's cues contribute to the predictive performance of all variables with $R^2 = .28$ for *performance impression*. An interesting result is that client's nonverbal cues explains variance in *performance impression* ($R^2 = .27$) almost as much as receptionist's own nonverbal cues. Similar results are observed for other skill impressions and are analogous to results reported in [40], where nonverbal cues of the interviewer contributed ($R^2 = 0.22$) to explaining the variance in applicant's hiring scores. The effect of gender on predictive power of nonverbal cues was investigated. The dataset was divided based on gender and regression experiments were rerun. No major difference in predictive performance of nonverbal cues was observed, with $R^2 = .28$ for female and $R^2 = .27$ for male participants for *performance impression*.

## 6.2 Performance Impressions & Personality Trait Impressions (RQ2)

To investigate the role of personality impressions predicting *performance impression*, a regression task was defined with the personality impressions as predictors. It is observed that the RF model explains up to 35% of the variance in the data. Similarly, these trait impressions performed moderately for other impressions, with highest $R^2$ achieved for *sociable* (0.43). Overall, all performance and skill impressions have moderate predictability [$R^2 \in (0.25, 0.43)$] using Big-5 impressions as predictors and finds support in the literature [7], answering RQ2 and validating the predictive power of Big-5 trait impressions in service encounters. As a next step, we combined the personality trait impressions and automatically extracted nonverbal cues to infer *performance impression*, and achieved $R^2 = .37$, which is marginally higher than each single source of information. In the case of *sociable* , up to 44% of variance in impression scores could be explained (highest among all skills). Overall, the highest performance for the inference task was achieved by combining nonverbal cues and Big-5 impressions implying that these impressions added extra information to the NVB cues.

## 6.3 Performance Impressions & Attractiveness (RQ3)

To further analyze the link between attractiveness variables and *performance impression*, a regression task was defined with the aim of evaluating the predictive power of attractiveness attributes as predictors. The ridge regression model performed best for this case and the results are presented in Table 7. From the table, it is observed that $R^2 < .20$, indicating that attractiveness variables had low predictive power. A similar observation was made while analyzing attractiveness attributes derived from still images in interview context by [40]. This result perhaps could be explained by the fact that raters annotated attractiveness variables by looking at still images rather than video clips. The methodology of using still images for attractiveness annotation, though has solid backing in psychology literature [4, 23], did not produce positive results in our case. In other works, authors reported a positive effect of physical attractiveness on performance impressions of sales representatives [1], teachers [4], and service workers [50]. This issue has to be investigated further.

In the computing literature using video instead of still images, Biel in [10] annotated two facets of physical attractiveness, and three facets of non-physical attractiveness on 442 1-min YouTube videos and reported that more attractive people were often judged as having more positive traits. Specifically, the authors reported correlation between *Agreeableness* and *Friendliness* to be $r = .57$ ($p < 0.001$). In this work, we found that the correlation between

**Table 7: Best inference performance results using NVB cues, personality traits (Big-5) impressions, Attractiveness impressions and various combination of impressions and NVB. All results were significant with $p < 0.05$ ($N = 169$). Best performing model is indicated by $^{*}$(RF); $^{**}$ (Ridge)**

| Impressions and Skills | | Baseline | | Nonverbal* | | Big-5* | | Attract** | | NVB + Big-5* | | NVB + Attract** | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | R2 | RMSE | R2 | RMSE | R2 | RMSE | R2 | RMSE | R2 | RMSE | R2 | RMSE |
| Performance | peImp | 0.0 | 0.0 | .30 | 1.31 | .35 | 1.22 | .18 | 1.30 | .37 | 1.18 | .21 | 1.31 |
| Professional | compe | 0.0 | 0.0 | .29 | 1.34 | .30 | 1.33 | .15 | 1.32 | .36 | 1.18 | .17 | 1.33 |
| | motiv | 0.0 | 0.0 | .30 | .88 | .29 | .89 | .14 | 1.02 | .34 | .83 | .12 | 1.03 |
| | satis | 0.0 | 0.0 | .29 | 1.42 | .32 | 1.36 | .16 | 1.31 | .36 | 1.28 | .13 | 1.30 |
| Social | intel | 0.0 | 0.0 | .26 | .81 | .27 | .82 | .13 | 1.08 | .28 | .80 | .11 | 1.07 |
| | posit | 0.0 | 0.0 | .32 | .79 | .33 | .78 | .12 | 1.06 | .41 | .69 | .17 | 1.06 |
| | socia | 0.0 | 0.0 | .33 | .85 | .43 | .73 | .13 | 1.07 | .44 | .71 | .12 | 1.07 |
| Communication | clear | 0.0 | 0.0 | .22 | 1.21 | .25 | 1.16 | .15 | 1.26 | .28 | 1.13 | .14 | 1.26 |
| | persu | 0.0 | 0.0 | .32 | 1.30 | .29 | 1.36 | .15 | 1.28 | .36 | 1.22 | .14 | 1.28 |

traits like *Agreeableness* (labeled on videos) and attractiveness variables like *Likeable* (labeled on images) to be very low ($r = .0.07$).

A hypothesis for this weak connection might be due to the difference in the amount of perceptual cues available for *performance impression* (video) and perceived attractiveness (still images). This hypothesis might find some support in [12], which reported that while a still image was a valid modality to infer various personality traits, a greater validity was achieved using audio-visual clips. The authors also state that "there are relations between physical attributes and personality traits, and subjects are quite aware of these relationships". For RQ3, we conclude that attractiveness variables inferred from still images have little connection to *performance impression* in our specific setting. In future work, we plan to investigate the use of video data for annotations of perceived attractiveness and its connections to performance and skill impressions.

## 7 CONCLUSION

This paper described our investigation of the interplay between nonverbal behavior cues, Big-5 personality trait, and attractiveness impressions in hospitality service encounters, a novel setting in multimodal interaction research. We extracted a number of relevant nonverbal cues automatically and studied their relationship with perceived performance and skill impressions.

Regarding RQ1, we found that receptionists who spoke faster, for greater duration, took longer turns, and had fewer silence events had high scores for performance and skill impressions. These results are supported by literature in marketing [51]. The inference task with NVB as predictors explains up to 30% of variance, and is comparable with results obtained for similar dyadic conversation settings in the literature.

Regarding RQ2, we found that Big Five personality trait impressions are predictors of performance and skill impressions. Specifically, receptionists who conveyed higher *Extraversion* were rated higher in terms of performance, while receptionists who were high in *Neuroticism* were rated lower with respect to performance. This is in line with work on Big-5 and job performance in psychology [7], and extends the findings to the hospitality reception desk scenario. An inference task with Big-5 impressions as predictors achieved a performance of $R^2 = 0.35$, while integrating NVB cues and Big-5 trait impressions results in slightly improved performance ($R^2 = 0.37$).

Regarding RQ3, our work found a negative correlation between attractiveness attributes like *likeable* and *friendly* and *performance impression* scores for women participants, while there was no correlation for men. Extending this further into a regression task using attractiveness attributes as predictors, we observed low predictive power ($R^2 < 0.18$) for all performance and skill impressions.

Finally, given the importance of service encounters on customer evaluation of quality of service, it seems essential for managers and employees to better understand how behavior might influence customer perception. Hence, this work could have implications for training and development of service employees. In the future, we plan to explore other behavioral cues including smiling, gaze, verbal content, and emotion recognition as features. We also plan to incorporate the findings of this work into a feedback system which provides automatic real-time feedback based on employee behavior.

## ACKNOWLEDGMENTS

This work was funded by the UBIMPRESSED project of the Sinergia interdisciplinary program of the Swiss National Science Foundation (SNSF). We would like to thank Denise Frauendorfer and the research assistants for their help with data collection, Yu Yu for providing the head nod features, Laurent Son Nguyen for discussions and the students of Vatel Switzerland for their participation, and Vatel's academic deans for their support.

## REFERENCES

[1] M. Ahearne, T. W. Gruen, and C. B. Jarvis. If looks could sell: Moderation and mediation of the attractiveness effect on salesperson performance. *Int. Journal of Research in Marketing*, 16(4):269–284, 1999.
[2] N. Ambady, M. A. Krabbenhoft, and D. Hogan. The 30-sec sale: Using thin-slice judgments to evaluate sales effectiveness. *Journal of Consumer Psychology*, 16(1):4–13, 2006.
[3] N. Ambady and R. Rosenthal. Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological bulletin*, 111(2), 1992.
[4] N. Ambady and R. Rosenthal. Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *Journal of personality and social psychology*, 64(3):431, 1993.
[5] N. Ambady and J. J. Skowronski. *First impressions*. Guilford Press, 2008.

[6] C. Barnum and N. Wolniansky. Taking cues from body language. *Management Review*, 78(6):59–61, 1989.

[7] 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.

[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. In *Proc. ACM ICMI*, 2011.

[9] J.-I. Biel and D. Gatica-Perez. The youtube lens: Crowdsourced personality impressions and audiovisual analysis of vlogs. *IEEE Trans. on Multimedia*, 15(1), 2013.

[10] J.-I. Biel and D. Gatica-Perez. Mining crowdsourced first impressions in online social video. *IEEE Transactions on Multimedia*, 16(7):2062–2074, 2014.

[11] S. Bonaccio, J. O'Reilly, S. L. O'Sullivan, and F. Chiocchio. Nonverbal behavior and communication in the workplace: A review and an agenda for research. *Journal of Management*, 42(5):1044–1074, 2016.

[12] P. Borkenau and A. Liebler. Trait inferences: Sources of validity at zero acquaintance. *Journal of personality and social psychology*, 62(4):645, 1992.

[13] L. Chen, G. Feng, C. W. Leong, B. Lehman, M. Martin-Raugh, H. Kell, C. M. Lee, and S.-Y. Yoon. Automated scoring of interview videos using doc2vec multimodal feature extraction paradigm. In *Proc. of ACM Int. Conf. on Multimodal Interaction*, pages 161–168. ACM, 2016.

[14] Y. Chen, Y. Yu, and J.-M. Odobez. Head nod detection from a full 3d model. In *Proc. IEEE ICCV Workshops*, 2015.

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

[16] T. DeGroot and S. J. Motowidlo. Why visual and vocal interview cues can affect interviewers' judgments and predict job performance. *Journal of Applied Psychology*, 84(6):986, 1999.

[17] M. Gabbott and G. Hogg. An empirical investigation of the impact of non-verbal communication on service evaluation. *European Journal of Marketing*, 34(3/4):384–398, 2000.

[18] M. Gabbott and G. Hogg. The role of non-verbal communication in service encounters: A conceptual framework. *Journal of Marketing Management*, 17(1-2):5–26, 2001.

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

[20] D. Gatica-Perez. Signal processing in the workplace. *IEEE Signal Process. Mag.*, 32(1), 2015.

[21] S. D. Gosling, P. J. Rentfrow, and W. B. Swann. A very brief measure of the big-five personality domains. *Journal of Research in personality*, 37(6):504–528, 2003.

[22] N. Guéguen and C. Jacob. Clothing color and tipping: Gentlemen patrons give more tips to waitresses with red clothes. *Journal of Hospitality & Tourism Research*, 38(2):275–280, 2014.

[23] D. S. Hamermesh and A. Parker. Beauty in the classroom: Instructors' pulchritude and putative pedagogical productivity. *Economics of Education Review*, 24(4):369–376, 2005.

[24] T. Hennig-Thurau, M. Groth, M. Paul, and D. D. Gremler. Are all smiles created equal? how emotional contagion and emotional labor affect service relationships. *Journal of Marketing*, 70(3):58–73, 2006.

[25] H. Hung and D. Gatica-Perez. Estimating cohesion in small groups using audio-visual nonverbal behavior. *IEEE Transactions on Multimedia*, 12(6):563–575, 2010.

[26] A. S. Imada and M. D. Hakel. Influence of nonverbal communication and rater proximity on impressions and decisions in simulated employment interviews. *J. Applied Psychology*, 62(3), 1977.

[27] C. Jacob, N. Guéguen, G. Boulbry, and R. Ardiccioni. Waitresse's facial cosmetics and tipping: A field experiment. *Int. journal of hospitality management*, 29(1):188–190, 2010.

[28] S. K. Johnson, K. E. Podratz, R. L. Dipboye, and E. Gibbons. Physical attractiveness biases in ratings of employment suitability: Tracking down the "beauty is beastly" effect. *The Journal of social psychology*, 150(3):301–318, 2010.

[29] H. S. Jung and H. H. Yoon. The effects of nonverbal communication of employees in the family restaurant upon customer's emotional responses and customer satisfaction. *Int. Journal of Hospitality Management*, 30(3):542–550, 2011.

[30] J. Kang and S. S. Hyun. Effective communication styles for the customer-oriented service employee: Inducing dedicational behaviors in luxury restaurant patrons. *Int. Journal of Hospitality Management*, 31(3):772–785, 2012.

[31] M. Knapp, J. Hall, and T. Horgan. *Nonverbal communication in human interaction*. Cengage Learning, 2013.

[32] M. Kuhn. A short introduction to the caret package. 2016.

[33] B. Lepri, N. Mana, A. Cappelletti, and F. Pianesi. Automatic prediction of individual performance from thin slices of social behavior. In *Proc. of ACM Int. Conf. on Multimedia*, pages 733–736. ACM, 2009.

[34] H.-F. Luoh and S.-H. Tsaur. Physical attractiveness stereotypes and service quality in customer–server encounters. *The Service Industries Journal*, 29(8):1093–1104, 2009.

[35] A. Madan, R. Caneel, and A. S. Pentland. Groupmedia: distributed multi-modal interfaces. In *Proc. of Int. Conf. on Multimodal Interfaces*, pages 309–316. ACM, 2004.

[36] MediaLabs. http://groupmedia.media.mit.edu/data.php.

[37] C. Mok, B. Sparks, and J. Kadampully. *Service quality management in hospitality, tourism, and leisure*. Routledge, 2013.

[38] S. Muralidhar, L. S. Nguyen, D. Frauendorfer, J.-M. Odobez, M. Schimd-Mast, and D. Gatica-Perez. Training on the Job: Behavioral Analysis of Job Interviews in Hospitality. In *Proc. 18th ACM Int. Conf. Multimodal Interact.*, pages 84–91, 2016.

[39] I. Naim, M. I. Tanveer, D. Gildea, and M. E. Hoque. Automated prediction and analysis of job interview performance: The role of what you say and how you say it. *Proc. IEEE FG*, 2015.

[40] L. S. Nguyen, D. Frauendorfer, M. S. Mast, and D. Gatica-Perez. Hire me: Computational inference of hirability in employment interviews based on nonverbal behavior. *IEEE Trans. on Multimedia*, 16(4), 2014.

[41] L. S. Nguyen and D. Gatica-Perez. I would hire you in a minute: Thin slices of nonverbal behavior in job interviews. In *Proc. of ACM Int. Conf. on Multimodal Interaction*, 2015.

[42] A. Pentland and T. Heibeck. *Honest signals: how they shape our world*. MIT press, 2010.

[43] F. Pianesi, N. Mana, A. Cappelletti, B. Lepri, and M. Zancanaro. Multimodal recognition of personality traits in social interactions. In *Proc. ACM ICMI*, 2008.

[44] B. Raducanu, J. Vitria, and D. Gatica-Perez. You are fired! nonverbal role analysis in competitive meetings. In *ICASSP 2009. IEEE Int. Conf. on*, pages 1949–1952. IEEE, 2009.

[45] T. C. Riniolo, K. C. Johnson, T. R. Sherman, and J. A. Misso. Hot or not: Do professors perceived as physically attractive receive higher student evaluations? *The Journal of General Psychology*, 133(1):19–35, 2006.

[46] R. T. Rust and R. L. Oliver. *Service quality: New directions in theory and practice*. Sage Publications, 1993.

[47] D. Sanchez-Cortes, O. Aran, M. S. Mast, and D. Gatica-Perez. A nonverbal behavior approach to identify emergent leaders in small groups. *Multimedia, IEEE Trans. on*, 14(3):816–832, 2012.

[48] B. R. Schlenker. *Impression management: The self-concept, social identity, and interpersonal relations*. Brooks/Cole Publishing Company Monterey, CA, 1980.

[49] P. E. Shrout and J. L. Fleiss. Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin*, 86(2), 1979.

[50] M. Söderlund and C.-R. Julander. Physical attractiveness of the service worker in the moment of truth and its effects on customer satisfaction. *Journal of Retailing and Consumer Services*, 16(3):216–226, 2009.

[51] D. Sundaram and C. Webster. The role of nonverbal communication in service encounters. *Journal of Services Marketing*, 14(5):378–391, 2000.

[52] C. Viswesvaran and D. S. Ones. Perspectives on models of job performance. *Int. Journal of Selection and Assessment*, 8(4):216–226, 2000.