Multimodal Analysis of Body Communication Cues in Employment Interviews

Laurent Son Nguyen
Idiap Research Institute and
EPFL
Switzerland
lnguyen@idiap.ch

Alvaro Marcos-Ramiro
University of Alcala
Spain
amarcos@depeca.uah.es

Marta Marrón Romera
University of Alcala
Spain
marta@depeca.uah.es

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

ABSTRACT

Hand gestures and body posture are intimately linked to speech as they are used to enrich the vocal content, and are therefore inherently multimodal. As an important part of nonverbal behavior, body communication carries relevant information that can reveal social constructs as diverse as personality, internal states, or job interview outcomes. In this work, we analyze body communication cues in real dyadic employment interviews, where the protagonists of the interaction are seated. We use a mixture of body communicative features based on manual annotations and automated extraction methods to successfully predict two key organizational constructs, namely personality and job interview ratings. Our work also confirms the multimodal nature of body communication and shows that the speaking status can be used to improve the prediction performance of personality and hirability.

Categories and Subject Descriptors

H.1.2 [**Information Systems**]: User/Machine Systems—*Human factors*

Keywords

Social computing, job interviews, body communication, multimodal interaction, personality, hirability

1. INTRODUCTION

Nonverbal communication plays a key role in face-to-face interactions as it conveys information in parallel with the spoken words, and becomes difficult to manipulate as it involves unconscious processes [18]. It has been shown to be a channel through which we reveal our internal state [28] or our personality traits [22, 12]; it therefore has a strong influence on how we are socially perceived. In nonverbal communication, body communication plays an important role. It comprises what the face, head, eyes, limbs, and trunk transmit. Although the importance of head gestures, facial expressions, and gaze has been demonstrated in the literature, we focus here on the analysis of body posture and gestures.

Gestures are an essential component of body communication as they are used to enrich the vocal content and aid listener comprehension by augmenting the attention, activating images or representations in the listener's mind, and increasing the recall of what is being said [18]. Moreover, restraining people from gesturing strongly affects the speakers' fluency [18]. Body posture is another important component of body communication; various emotions such as fear, sadness, or happiness have been shown to be correctly inferred from a person's pose [18]. In conversations, body posture can be used as markers during a conversation: for instance, changes of body posture can precede a long utterance and may be kept for the duration of the speaking turn [18]. In this sense, both gestures and postures are inherently multimodal, in that they do not only occur in the visual modality, but are conditioned on the speaking status (*i.e.*, audio modality) of the person. For this reason, we believe that it is necessary to consider the speaking status when analyzing posture and gestures.

Used in nearly every organization, the employment interview is an interpersonal interaction between at least one interviewer and a job applicant, where the latter is evaluated for an open position. It aims to assess the candidate's suitability for the job at hand and is one of the most popular tools for this task [27]. Interviews are inherently social as they require face-to-face interaction among the protagonists [16]. Since applicants and recruiters typically meet for the first time, employment interviews are a form of *zero-acquaintance* interactions [2]. Apart from the résumés, the applicant's verbal and nonverbal behavior is sometimes the only information that recruiters have to forge an opinion.

In this study, we address two main research questions: First, we investigate whether job hirability impressions and self-rated personality can be predicted using body commu-

ICMI'13, December 9–13, 2013, Sydney, Australia
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-2129-7/13/12 ...$15.00.
http://dx.doi.org/10.1145/2522848.2522860.

nication cues; second, leveraging on the multimodal nature of posture and gestures, we examine whether the knowledge of the speaking status can be used to improve the prediction of personality and hirability.

To address these research questions, several tasks were defined. First, we used a dataset of real job interviews, including self-reported personality scores from questionnaire data and expert-rated hirability impressions. Then, we extracted a rich mixture of body cues from both manual annotations and automated extraction methods, including the speaking status in the extraction process. As a next step, we analyzed the differences of body communication cues depending on whether the applicant was speaking or silent. Last, we evaluated the predictive validity of the extracted nonverbal cues with respect to hirability impressions and self-rated personality using a regression task.

The main contributions of this paper are therefore: (1) The prediction of two organizational constructs in job interviews, namely personality and hirability, using body nonverbal cues. To our knowledge, [24] is the first work in systematically analyzing audio-visual nonverbal behavior in employment interviews. In this present work, we extend the study in [24] by predicting personality traits in addition to hirability, and systematically focus on postures and gestures. (2) The systematic analysis of body communication cues for the prediction of social constructs. (3) The exploitation of the multimodal nature of body communication to improve the prediction performance of personality and hirability.

The following section presents the related work studying the relationships between nonverbal behavior, personality, and hirability, from both the psychology and the computing perspectives. In Section 3, we present the dataset of real job interviews used in this study. In Section 4, we present the body communication cues extracted from manual annotations and automatic hand activity images. We then present our framework for predicting the social variables in Section 5, present and discuss the results in Section 6, and finally conclude in Section 7.

2. RELATED WORK

2.1 Related work in psychology

In job interviews, applicant nonverbal behavior has a remarkable impact on the hiring decision. For instance, research shows that applicants who use more immediacy nonverbal behavior (*i.e.*, eye contact, smiling, body orientation toward interviewer, less personal distance) are perceived as being more hirable, more competent, more motivated, and more successful than applicants who do not [17]. Organizational psychology literature suggests that the relation between nonverbal behavior and job interview outcomes can be based on the immediacy hypothesis, which claims that the applicant reveals through his or her immediacy behavior a greater perceptual availability between the applicant and the recruiter. This in turn leads to a positive affect in the recruiter and therefore to a more favorable evaluation [17].

The five-factor structure, commonly known as the Big-Five, has received extensive support in psychology for describing personality [15]. This framework is a hierarchical model of personality traits with five broad factors, which represent personality at its highest level of abstraction [15]. The model suggests that most individual differences in human personality can be classified into five empirically-derived bipolar factors, namely extraversion, agreeableness, conscientiousness, neuroticism, and openness to experience (see Table 1). Previous studies have shown the importance of personality in employment interviews. In particular, conscientiousness and extraversion correlate significantly with employability ratings for conventional and enterprising jobs [9]. The validity of these inferences have been demonstrated, as conscientiousness is significantly correlated with job performance across all job occupations and extraversion is positively related to occupations requiring social skills [5].

Table 1: Big-Five traits and related adjectives [15]

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

Body communication plays an important role in conjunction with spoken words to enhance the communication in face-to-face interactions [18], which is also the case in employment interviews. Highly employable job candidates usually produce more hand gestures, both in frequency and amplitude [3, 12], and have the tendency to lean forward more [14]. These kinesic nonverbal cues were also found to be associated with some personality dimensions. For instance, rapid body movement is shown to be related to extraversion and agreeableness, while relaxed body posture has been associated with conscientiousness [8].

2.2 Related work in computing

The advent of cheap sensors, in combination with improved automated perceptual methods, have enabled the development of computational methods to predict social constructs. As a key construct to explain inter-individual differences, Big-Five personality has been studied in various settings. These include small group interactions [25], video blogs [7], or human-computer interaction [6]. While most existing studies rely on audio nonverbal cues (prosody, speaking-turn-based features) and visual cues (head nods, visual activity, head pose) to predict personality, few studies have investigated the use of body communication. The approach in [6] used manually annotated hand movements and posture to predict personality in a human-computer interaction. Other human-computer interaction studies [4, 23] have examined the link between specific personality dimensions (extraversion, friendliness, dominance) and body posture for enabling embodied conversational agents with the ability to produce postures given the personality tendency. To our knowledge, no computational study has specifically investigated the role of posture and gestures for the prediction of personality. Moreover, the speaking status has not been taken into account to jointly analyze these cues.

Despite the ubiquity and the importance of employment interviews, very few computational studies have examined the role of behavior in such settings. To our knowledge, two studies have approached this scenario. The work in [11] studied the relation between nonverbal behavior and outcomes in a simulated dyadic negotiation configuration. The approach in [6] addressed short self-presentations in a human-computer interaction context, which resembles the job interview setting. Previous work by ourselves and collaborators have studied the employment interview setting in two ways. First in [19], a study on recognizing stress levels from acoustic features was done using 14 interviews, but no connection with hirability were explored. Second in [24], we presented the first study on hirability prediction based on nonverbal cues. Here, we extend this analysis by focusing on body communication cues, and by adding the task of predicting self-rated personality traits.

Relatively few automatic methods exist to extract body pose from video with conversational constructs in mind. In [20], hands location and speed, together with an approximate body pose were used to build an activity descriptor. Generic body pose retrieval has been the subject of enormous attention. Famously, [26] proposed the use of a classifier in RGB-D images, resulting in the commercial Microsoft Kinect device. As a proxy for body activity, some methods like [7] measure basic image parameters, given a coarse estimation. The subject has also been approached by the wearable computing field [13]. However, the needed sensors can condition the naturality of the body movements.

3. DATASET

3.1 Participants and scenario

We used a dataset of 43 real employment interviews, originally described in [24]. The four-hour job at stake consisted in recruiting people on the street for psychology experiments, and was remunerated with 200 Swiss Francs. In order to gather subjects for the study, we advertised the job in three different Swiss universities using multiple communication channels. Due to the large participation of students, the average age was 24 years, with a standard deviation of 5.7 years.

Before starting the interview, applicants were asked to fill in a consent form where they accepted that the interview would be audio- and video-recorded, and that the data could be used for research purposes. They then completed a questionnaire to assess their Big-Five personality scores.

For the interview itself we used a structured design, where the sequence of instructions and questions remained constant across interviews in order to ensure that comparisons could be made between job candidates. The job applicants were asked to answer four behavioral questions related to past experiences in specific situations requiring specific social skills, namely cases (1) where communication skills were required, (2) where persuasion skills were required, (3) of conscientious/serious work, and (4) where stress was properly managed. The interviews were dyadic, and the recruiter was facing the job applicant. The protagonists were seated at both sides of a table (see Figure 1). In total, the dataset comprises over 475 minutes of recordings, with an average interview duration of $\sim$ 11 minutes.

3.2 Technical setup

Audio and standard color video were recorded for both the applicant and the interviewer. Audio was registered by a Microcone [1] microphone array, which automatically segments speaker turns, while recording at 48kHz. The video was recorded at 30 frames per second with two VGA cameras. Audio-video synchronization was processed manually by adjusting the delay between the sound and the lip movements. An illustration of the interview setup is illustrated in Figure 1.

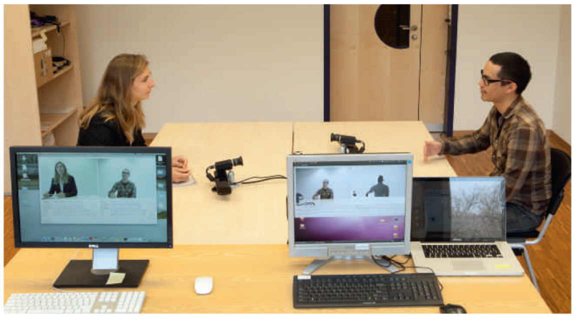

Figure 1: Setting and data collection. Audio and video for interviewer and participant were recorded and synchronized.

3.3 Social variables

The Big-Five personality variables were assessed using the standard NEO-FFI-R questionnaire [10], which is formed by 60 items (12 per personality dimension). Hirability scores were manually coded by a task-trained M.S. student in organizational psychology. The annotations were completed after watching the full audio-video recording of the interview, including both the recruiter and the job candidate. Four hirability scores were coded based on the answers to the four behavioral questions: communication, persuasion, conscience, and stress resistance. A score between 1 (very low) and 5 (very high) was assigned for each variable. In addition to these four scores, a general score (hiring decision) was given based on the general impression made by the applicant, ranging from 1 to 10. In order to validate the reliability of the annotator, a second expert coder rated 10 job interviews. The inter-rater agreement was satisfactory, with Pearson's correlation coefficient ranging from 0.69 for conscience to 0.99 for persuasion.

4. FEATURE EXTRACTION

Our objective is to explore the use of multimodal body communication cues to predict hirability and personality. To this end, we leverage on automatic speaker segmentations provided by the Microcone device, manual annotations of postures and gestures, and automatic extraction of hand movement and hand activity zones. We present the method used to extract detailed gesture- and posture-based nonverbal cues.

4.1 Annotations of body activity

Five classes were defined based on the occurrences in the dataset and the relevance in the nonverbal communication literature [18]: hidden hands, hands on table, gestures on table, gestures, and self-touch. They constitute an approximation for the applicant's body posture and gestures. Applicants were seated, therefore the posture was for a large part defined by the position of their arms. Other posture classes such as leaning forward or backward were also considered, but were discarded as the observed variability of such postures was low. Class definitions and examples can be seen in Table 2 and Figure 2, respectively.

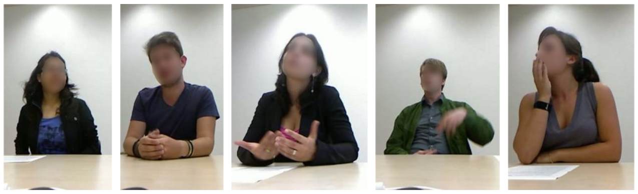

Figure 2: Class examples. From left to right: hidden hands, hands on table, gestures on table, gestures, self touch.

Table 2: Class descriptions.

Class	Description
Hidden hands (HH)	No hands visible in the image
Hands on table (HT)	Resting the hands in the table
Gestures on table (GT)	Gesturing while the hands are close to the table, or the arms resting on it
Gestures (G)	Gesturing while the hands are not close to the table
Self-touch (ST)	Touching face, hair or torso with one or both hands

Applicant body activity was annotated at the frame level by one person, with the help of a purpose-built script. To reduce the amount of frames to label, annotations were made every 15 frames (0.5 seconds); this temporal resolution was sufficient as no missing labels were observed while playing the full video at regular speed. To further reduce the amount of frames to label, we applied a motion threshold to the videos and annotated frames only when sufficient movement was present; unannotated frames in-between were assigned the same label as the latest annotated frame. This procedure allowed us to reduce the number of frames to annotate by 35%. In total, over 23000 frames were labeled. In order to assess the reliability of the annotations, a second person annotated 63 minutes of the dataset (≈5000 frames), and interrater agreement was satisfactory (using Cohen's Kappa: $\kappa = 0.81$).

4.2 Automatic features

4.2.1 *Speaking status*

Speaking status was extracted using the Microcone, which automatically segments speaker turns by using a filter-sum beamformer followed by a post-filtering stage in each of the spatial segments of the microphone array. The resulting speaker segmentations were stored in a file containing the relative time (start and end) and the speaker identifier. The objective performance of the speaker segmentation was not evaluated, but we manually inspected all segmentation files and observed only a small number of segmentation errors, even for short segments or overlapping speech.

4.2.2 *Hand speed*

To obtain estimates of hand speed, we used the method presented in [20] to compute the hand likelihood map for each frame of a video. This method makes the assumptions that the hands are the quickest parts of the video, that they are not the face, and that they have skin color. Based on these assumptions, the hand likelihood map can be computed as the product of the dense optical flow map, the binary face-mask image, and the skin-color segmentation binary image. Because the amount of data processing was substantially larger than in [20], we implemented a simple but effective method to obtain the hand speed image: we multiplied the hand likelihood map with the pixel frame-difference and normalized it by the distance between the head and the table to account for variations in the camera placement. An illustration of the procedure to compute the hand speed image is displayed in Figure 3. As a last step, we obtained the hand speed energy $e_H(t)$ by summing all pixels of the hand speed image, resulting in a single value for the hand speed estimate for each frame of a video.

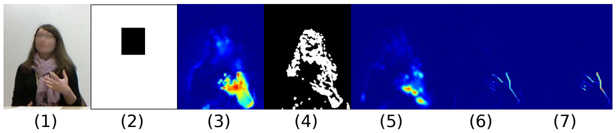

Figure 3: Illustration of the hand speed image computation: (1) original image, (2) face mask, (3) optical flow map, (4) skin-color segmentation image, (5) hand likelihood map, (6) frame difference, and (7) resulting hand speed image.

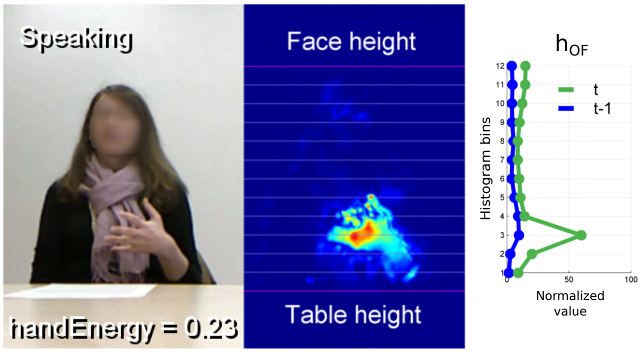

Figure 4: Left: input image with overlaid speaking status and hand energy value. Center: Dense optical flow and image height division. Right: Activity histograms $\mathbf{h}_{OF}$. Best viewed in color.

4.2.3 *Image activity histograms*

In order to obtain information about the hand position of the participants, we created an image activity descriptor along the vertical axis of the image. We defined a 12-bin histogram $\mathbf{h}_{OF}(t)$, which accumulates energy in different height bands of the dense optical flow image (normalized by the distance between the table and the head). The histogram is able to capture two important factors which condition the applicant's visual activity, *i.e.* hand speed and hand height, which makes this feature suitable for the analysis of seated participants. Moreover, as the method is based on dense optical flow, it is appearance invariant, which makes it suitable for the analysis of subjects with different skin colors. An illustration of the image activity histogram can be seen in Figure 4.

4.3 Nonverbal cue encoding

Next, we describe the method to encode the nonverbal cues from the manual annotations of body activity, the speaker segmentations, the hand speed estimates, and the image ac-

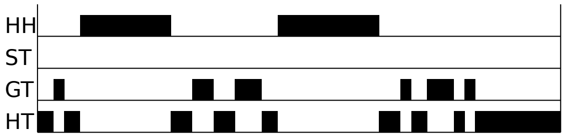

Figure 5: Illustration for cues based on annotations of body activity. There are two HH events, six GT events, nine HT events, and no ST events. Statistics are computed from event durations. If no event occurred, the statistics are set to zero.

tivity histograms. These cues will then be used as features for the prediction of personality and hirability. approach to obtain these cues.

4.3.1 Cues based on annotations of body activity

Nonverbal cues were extracted from the manual annotations of body activity. To capture a ”big picture” of the body activity, they were based on rich statistics derived from event durations. Events were defined as a sequence of frames where the applicant showed the same type of body activity, and are characterized by their starting time and duration (see Figure 5). For all the activity classes, we computed the number of events, mean, median, standard deviation, lower and upper quartiles, minimum, maximum, range, position (in time) of shortest and longest events, and total relative time. It should be noted that it was possible for a given class to be missing in a given sequence. We addressed this by introducing a binary variable indicating whether at least an event occurred or not. The statistics on turn durations were set to zero if no event occurred. The list of body communication cues based on manual annotations is included in Table 3.

4.3.2 Cues based on hand speed and activity histograms

The hand speed approximation $e_H(t)$ and the image activity histograms $\mathbf{h}_{OF}(t)$ (sections 4.2.2 and 4.2.3) not only provide information on *how much* hand movement occurred at a given instant, but also on *where* these hand movements occurred. We also extracted nonverbal cues based on those activity descriptors. To account for short bursts of hand movement characterized by quick changes of hand speed (which could be associated with beat gestures), we computed the hand acceleration. We defined the global hand acceleration at time t as $a_H(t) = |e_H(t) - e_H(t-1)|$, and the image-height-dependent acceleration as $\mathbf{a}_{OF}(t) = |\mathbf{h}_{OF}(t) - \mathbf{h}_{OF}(t-1)|$.

To extract nonverbal cues from the univariate time series e_H and a_H, we computed the mean, median, standard deviation, minimum, maximum, range, quartiles, proportion of non-zero elements, and zero-crossing rate. We also computed statistics related to the histogram main mode (*i.e.*, the position of the maximum histogram bin) to account for hand position: mean, median, standard deviation, quartiles, and zero-crossing rate. Table 3 shows the list of automatic body communication cues used in this study.

4.3.3 Exploiting the speaking status

To exploit the finding in psychology stating that body communication is conditioned on the speaking status [21, 18], we computed the statistics on manual body activity event durations, hand speed and acceleration, and activity and acceleration histogram modes for four different cases: (1) the unimodal case, *i.e.* without taking into account the speaking status, (2) the speaking case, *i.e.* only using frames for which the applicant was speaking, (3) the silent case, and (4) the aggregated case, *i.e.* aggregating the three previous cases. Table 3 shows the list of all the body communication cues used in our study.

Table 3: List of the manual and automatic nonverbal cues used in this study. Each statistical cue was computed for the unimodal (*i.e.* not taking the speaking status into account), speaking, silent, and aggregated cases (*i.e.* aggregating unimodal, speaking, and silent).

Manual features:		
Posture class	**Statistics**	**Speaking status**
Hidden hands (HH) Self-touch (ST) Hands on table (HT) Gestures on table (GT) Gestures (G)	mean, median, std, quartiles, # of events, min., max., range, rel. time, pos. of min./max., exists	Unimodal, Speaking, Silent, Aggregated

Automatic features:		
Time-series	**Statistics**	**Speaking status**
Hand velocity (HV) Hand acceleration (HA)	mean, median, std, quartiles, zero-crossing rate, min., max., range, prop. non-zero	Unimodal, Speaking, Silent, Aggregated
Histograms (mode)	**Statistics**	**Speaking status**
Hand velocity histogram (HVH) Hand acceleration histogram (HAH)	mean, median, std, quartiles, zero-crossing rate,	Unimodal, Speaking, Silent, Aggregated

5. PREDICTION OF TASKS

In order to analyze the predictive validity of body posture with respect to self-rated personality traits and hirability impressions, we defined a regression problem which aims at predicting the exact hirability and personality scores, where each social variable is considered as an independent regression task. To this end, we used a leave-one-interview-out cross validation strategy. Two regression methods were used for predicting personality and hirability. We used ridge regression as the first prediction model, a linear model where the parameters are estimated by minimizing the sum of squared errors plus an l_2 regularization term (referred as the ridge parameter), which prevents the model from overfitting. The ridge parameter was estimated automatically using 10-fold inner cross-validation. As a second prediction method, we used random forest with 1000 trees, which is a robust non-linear regression model.

Given the large number of features ($D > 300$) compared to the number of data points ($N = 43$), we decided to analyze sub-groups of features independently. This allowed the regression model to be correctly learned, and enabled the analysis of the predictive validity of specific postures and speaking cases. For the nonverbal features based on the manual annotations, five feature groups were defined based on the annotated body activity classes (described in Table 2). For the automatic cues, we used the hand movement e_H,

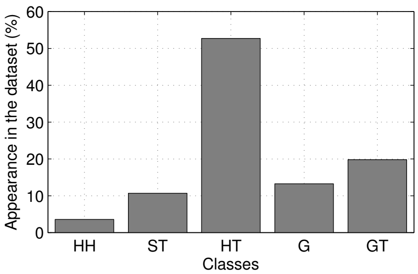

Figure 6: Frequency of each class in the dataset.

the hand acceleration a_H, the activity histogram $\mathbf{h}_{OF}$, and the acceleration histogram $\mathbf{a}_{OF}$ cues as four feature groups.

In order to test whether exploiting the speaking status improves the prediction accuracy, we further segmented the feature groups into four sub-groups: (1) unimodal features, *i.e.* obtained without taking into account the speaking status, (2) silent features only, (3) speaking features only, and (4) aggregated features, *i.e.* the concatenation of unimodal, silent, and speaking cues. Prediction results using specific posture cues, speaking status, and regression methods are reported and discussed in Section 6.3.

6. RESULTS AND DISCUSSION

6.1 Annotation statistics

In Table 4, we show descriptive statistics of the personality and hirability variables used in this study. We observe that except communication and conscience, all hirability measures were significantly correlated with each other. Also, extraversion was found to be significantly and positively correlated with three hirability scores: hiring decision, conscience, and stress resistance. This suggests that extraverts were seen as more employable by the coder. This finding is supported by the related psychology literature, which finds extraversion as a valid predictor of performance in jobs characterized by a high level of social interactions [5], as it is the case here.

The class distribution of the corpus is shown in Figure 6. We observe that *hands on table* accounted for more than half of the labels. The dataset was recorded in a real setting, therefore it reflects the natural tendency of the participants while being seated. It should be noted that in 34.2% of the data the subject was silent while listening to the interviewer. Our proxy for beat gestures (*gestures* and *gestures on table*) were present 33.6% of the time. The least represented class was *hidden hands*, while *self-touch* appeared almost as often as *gestures*.

6.2 Analysis of speaking status

In order to test whether our initial assumption stating that body communication was conditioned on the applicant's speaking status, we computed the Student's t-test to examine whether some significant differences in feature values between speaking and silent existed. In Table 5, we display the significantly different features ($p < 0.05$), and report whether the larger value was associated with moments when the job applicant was silent or speaking.

We observe that job applicants gestured more when they were speaking (more *gestures* and *gestures on table* time, longer events, larger range of durations; larger hand speeds; larger hand accelerations). Inversely, interviewees self-touched and kept their hands on the table longer when listening to the interviewer. This findings validate our main assumption of the multimodal nature of hand gestures and body posture, based on the nonverbal communication literature [18, 21]. Furthermore, we observe that the automatic features based on hand speed and hand acceleration were also conditioned on the speaking status.

Table 5: Feature significantly different (p < .05) between speaking and silent, using Student's t-test.

Feature group	Larger feature value for silent	Larger feature value for speaking
Hid. hnds		# of events
Slf-Tch	rel. time, median, max., min., quartiles, range	# of events
Hnds table	rel. time, mean, std, upper quartile	# of events
Gestures		rel. time, mean, median, std, max., upper quartile, range, exist, # of events
Gest. table		rel. time, mean, median,std,max., min., range, quartiles, # of events, exist
Hnd speed	min, zero-crossing rate	mean, median, max., quartiles, range, non-zero prop.
Hnd acceleration		mean, median, max., min., quartiles, non-zero prop.

6.3 Prediction of hirability and personality

One of the research questions of this study was to investigate whether hirability and personality could be inferred using body communication cues as predictors. In order to evaluate the prediction accuracy of our method, we used the standard coefficient of determination R^2, which can be seen as the amount of variance explained by the evaluated model. In Table 6, we report the results for which the R^2 values were higher than 0.1. From those findings, several observations can be made.

Except for communication ability, all hirability scores were inferred above the R^2 threshold using multimodal body communication cues. Importantly, we achieved $R^2 = 0.209$ for the hiring decision score using automatically extracted activity histogram features (aggregation of unimodal, speaking, and silent) and ridge regression as a prediction method. This finding demonstrates the potential of predicting job interview outcomes using body communication cues.

For personality, we show that prediction can be done using body communication cues only, which was to our knowledge not analyzed systematically prior to this work. Using such nonverbal features showed prediction performance comparable to the related work in social computing. Extraversion prediction ($R^2 = 0.165$) was found to be less accurate than in the state of the art (*e.g.* [7]), but results obtained for openness to experience ($R^2 = 0.238$), agreeableness ($R^2 = 0.140$), and conscientiousness ($R^2 = 0.165$) were positive.

Table 4: Descriptive statistics (mean, std, and Pearson's correlation) of personality and hirability ($^{*}p < .05$, $^{\dagger}p < .005$)

	μ	σ	1	2	3	4	5	6	7	8	9	10
1. HirDecision	6.209	1.753		0.536†	0.770†	0.662†	0.707†	0.501†	0.014	-0.199	0.136	0.119
2. Communication	2.953	0.872			0.429†	0.268	0.306*	0.279	-0.113	-0.080	0.037	0.080
3. Persuasion	2.977	1.080				0.493†	0.520†	0.250	0.076	-0.210	0.059	0.113
4. Conscience	3.070	1.033					0.602†	0.358*	0.070	-0.235	-0.003	0.289
5. StressRes	3.047	0.722						0.341*	0.124	-0.182	0.035	0.263
6. Extraversion	4.008	0.434							0.037	-0.292	0.449†	0.313*
7. Openness	3.736	0.527								-0.049	0.112	-0.072
8. Neuroticism	2.210	0.573									-0.168	-0.578†
9. Agreeableness	4.144	0.418										0.346*
10. Conscientiousness	4.106	0.640										

Table 6: Prediction results for hirability impressions (1-5) and self-rated personality (6-10) using manual (M) and automatic (A) cues. $\mathbf{R^2}$ was used to evaluate the prediction performance. Only results with $\mathbf{R^2 > 0.1}$ are reported.

Hirab. variable	Feature group	Spking status	Regr. method	$\mathbf{R^2}$
1. HirDec	Gesturing (M)	Silent	Ridge	0.196
	Activ. hist. (A)	Silent	Ridge	0.177
	Activ. hist. (A)	Silent	RF	0.141
	Activ. hist. (A)	Aggr.	Ridge	**0.209**
	Activ. hist. (A)	Aggr.	RF	0.139
	Acc. hist. (A)	Silent	Ridge	0.180
	Acc. hist. (A)	Silent	RF	0.114
2. Comm	-	-	-	-
3. Consc	Hid. hnds (M)	Speak	RF	0.174
	Hid. hnds (M)	Aggr.	RF	0.103
	Gest. table (M)	Speak	RF	0.159
	Gest. table (M)	Aggr.	Ridge	0.150
	Gest. table (M)	Aggr.	RF	**0.200**
	Activ. hist. (A)	Silent	RF	0.119
	Activ. hist. (A)	Aggr.	RF	0.109
4. Persuas	Hid. hnds (M)	Aggr.	Ridge	**0.235**
	Activ. hist. (A)	Silent	RF	0.204
	Activ. hist. (A)	Aggr.	RF	0.109
	Activ. hist. (A)	Aggr.	Ridge	0.127
	Acc. hist. (A)	Silent	Ridge	0.144
	Acc. hist. (A)	Silent	Ridge	0.109
5. StrRes	Activ. hist. (A)	Aggr.	RF	**0.103**

Pers. variable	Feature group	Spking status	Regr. method	$\mathbf{R^2}$
6. Extra	Hid. hnds (M)	Unimod.	RF	**0.165**
	Hid. hnds (M)	Speak	RF	0.124
	Hid. hnds (M)	Aggr.	Ridge	0.154
	Hid. hnds (M)	Aggr.	RF	0.139
	Slf-tch (M)	Unimod.	RF	0.112
	Slf-tch (M)	Aggr.	RF	0.127
	Gest. table (M)	Unimod.	RF	0.152
	Gest. table (M)	Speak	RF	0.137
7. Open	Slf-tch (M)	Unimod.	Ridge	0.109
	Slf-tch (M)	Silent	RF	0.103
	Hnds table (M)	Silent	RF	**0.238**
	Hnds table (M)	Silent	RF	0.177
8. Neuro	-	-	-	-
9. Agree	Hid. hnds (M)	Speak	RF	**0.140**
	Gest. table (M)	Speak	RF	0.111
10. Consc	Hid. hnds (M)	Unimod.	Ridge	0.136
	Hid. hnds (M)	Silent	Ridge	**0.165**
	Gest. table (M)	Unimod.	Ridge	0.136
	Gest. table (M)	Silent	Ridge	**0.165**

Only six prediction scores where $R^2 > 0.1$ were achieved using unimodal features (*i.e.* without taking into account the speaking status of the job applicant when extracting the cues). In comparison, 33 prediction scores were achieved by using body communication cues conditioned on the speaking status. This finding further shows the intimate link between speaking status and body communication in this job interview setting. Furthermore, we show that leveraging on this finding can improve the prediction of social constructs.

We observe that automatic hand activity cues were predictors of hirability ratings. Indeed, the best prediction results for the hiring decision and stress resistance were achieved using automatic cues based on activity histograms. For the hirability variables of persuasion and conscience, the use of automatic features decreased the prediction accuracy compared to manual features (from 0.235 to 0.205 and 0.200 to 0.119, respectively). This finding suggests that manual annotations of postures might not be necessary, depending on the social construct of interest. The use of automatic body communication cues was however found to show poor performance for self-rated personality traits.

7. CONCLUSION

We conducted a systematic study on applicant body communication in job interviews with respect to hirability impressions and self-rated personality. Additionally, we leveraged on findings in psychology suggesting a strong link between body communication and speech to analyze body communication from a multimodal perspective.

We used a dataset of 43 real job interviews. We extracted a rich mixture of body communication features from manual annotations of body activity and automatic hand speed descriptors. To account for the speaking status, these features were conditioned on whether the applicant was silent or speaking. By analyzing the corresponding differences in feature values, we validated our main assumption stating that speaking and silent differences existed.

As reported in Section 6, we show that the prediction of interview outcomes using body communication cues is a promising task. To our knowledge, the only work systematically analyzing employment interviews is our previous work [24]; we have contributed new findings for the case when only body communication cues are used, as opposed to other nonverbal cues such as speaking activity, head gestures, or prosody. We also show that body communication cues can be used to predict applicant personality traits, achieving results similar to the state of the art. The reported results also demonstrate that exploiting the intimate link between body communication and speaking status helps the inference of personality and hirability.

The prediction of some of the constructs analyzed in this work rely on manual annotations of body activity. This is the case of personality traits, where no automatic feature could produce accurate prediction scores. However, results show that in certain cases, using the automatic hand speed estimates yielded higher prediction results than manual features. This is the case for the variable of hiring decision. This finding underlines the relevance of automatic hand speed estimates for the analysis of employment interviews, even if these estimates are coarse.

Whereas in this present work we limited our analysis to applicant behavior, we plan for future work to analyze the other half of the dyad, *i.e.* the interviewer. Such an analysis could allow us to determine whether the interviewer's behavior could be used to predict interview outcomes or provide information about the applicant's personality.

Acknowledgments

This work was funded by the SONVB project of the Sinergia interdisciplinary program of the Swiss National Science Foundation (SNSF), the VISNU project of the Spanish Ministry of Science and Innovation, and the FPI program of the University of Alcala. Special thanks to Marianne Schmid-Mast and Denise Frauendorfer (Université de Neuchâtel) for the collaboration on experimental design and data collection. We would also like to thank Daniel Pizarro-Perez (Université d'Auvergne) for the fruitful discussions, Clémentine Duc and Amandine Loude (Université de Neuchâtel) for the annotations, and the applicants who kindly accepted to share their data.

8. REFERENCES

[1] Microcone: intelligent microphone array for groups [online]. Available: http://www.dev-audio.com/products/microcone/.

[2] N. Ambady, M. Hallahan, and R. Rosenthal. On judging and being judged accurately in zero-acquaintance situations. *Journal of Personality and ...*, 69(3):518–529, 1995.

[3] N. Anderson and V. Shackleton. Decision making in the graduate selection interview: A field study. *Occupational Psychology*, 63(1):63–76, 1990.

[4] G. Ball and J. Breese. Relating Personality and Behavior : Posture and Gestures. *Lecture Notes in Computer Science*, 1814:196–203, 2000.

[5] M. Barrick and M. Mount. The big five personality dimensions and job performance: a meta-analysis. *Personnel Psychology*, 44(1):1–26, 1991.

[6] L. Batrinca, N. Mana, B. Lepri, F. Pianesi, and N. Sebe. Please, tell me about yourself: automatic personality assessment using short self-presentations. *Proc. Int. Conf. on Multimodal Interactions*, pages 255–262, 2011.

[7] J.-I. Biel and D. Gatica-Perez. The YouTube Lens: Crowdsourced Personality Impressions and Audiovisual Analysis of Vlogs. *IEEE Transactions on Multimedia*, 15(1):41–55, Jan. 2013.

[8] J. Burnett and S. Motowidlo. Relations between different sources of information in the structured selection interview. *Personnel Psychology*, 51(4):963–983, 1998.

[9] M. Cole, H. Feild, and W. Giles. Job type and recruiters' inferences of applicant personality drawn from resume biodata: Their relationships with hiring recommendations. *Journal of Selection and Assessment*, 12(4):363–367, 2004.

[10] P. T. Costa and R. R. McCrae. *Neo PI-R Professional Manual.* Psychological Assessment Resources, 1992.

[11] J. Curhan and A. Pentland. Thin slices of negotiation: Predicting outcomes from conversational dynamics within the first 5 minutes. *Applied Psychology*, 92(3):802, 2007.

[12] T. DeGroot and J. Gooty. Can Nonverbal Cues be Used to Make Meaningful Personality Attributions in Employment Interviews? *Journal of Business and Psychology*, 24(2):179–192, Feb. 2009.

[13] S. Feese, B. Arnrich, G. Troster, B. Meyer, and K. Jonas. Detecting posture mirroring in social interactions with wearable sensors. *Wearable Computers, IEEE International Symposium*, 2011.

[14] R. Gifford, C. F. Ng, and M. Wilkinson. Nonverbal cues in the employment interview: Links between applicant qualities and interviewer judgments. *Applied Psychology*, 70(4):729–736, 1985.

[15] S. Gosling. A very brief measure of the Big-Five personality domains. *Research in Personality*, 37(6):504–528, Dec. 2003.

[16] J. Howard and G. Ferris. The Employment Interview Context: Social and Situational Influences on Interviewer Decisions. *Journal of Applied Social Psychology*, 26(2):112–136, 1996.

[17] A. S. Imada and M. D. Hakel. Influence of Nonverbal Communication and Rater Proximity on Impressions and Decisions in Simulated Employment Interviews responsibilities that produce different re-. *Journal of Applied Psychology*, 62(3):295–300, 1977.

[18] M. L. Knapp and J. A. Hall. *Nonverbal communication in human interaction.* Wadsworth, Cengage Learning, 7 edition, 2009.

[19] H. Lu, M. Rabbi, G. T. Chittaranjan, D. Frauendorfer, M. Schmid Mast, A. T. Campbell, D. Gatica-Perez, and T. Choudhury. StressSense: Detecting stress in unconstrained acoustic environments using smartphones. In *Proc. Int. Conf. on Ubiquitous Computing*, 2012.

[20] A. Marcos-Ramiro, D. Pizarro-Perez, M. Marron-Romera, L. Nguyen, and D. Gatica-Perez. Body communicative cue extraction for conversational analysis. In *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*, 2013.

[21] D. McNeill. So You Think Gestures Are Nonverbal? *Psychological Review*, 92(3):350–371, 1985.

[22] L. P. Naumann, S. Vazire, P. J. Rentfrow, and S. D. Gosling. Personality judgments based on physical appearance. *Personality & social psychology bulletin*, 35(12):1661–71, Dec. 2009.

[23] M. Neff, Y. Wang, R. Abbott, and M. Walker. Evaluating the effect of gesture and language on personality perception in conversational agents. *Intelligent Virtual Agents*, 2010.

[24] L. Nguyen, D. Frauendorfer, M. Schmid Mast, and D. Gatica-Perez. Hire me: Computational inference of hirability in employment interviews based on nonverbal behavior. Technical report, Idiap Research Institute.

[25] F. Pianesi, N. Mana, A. Cappelletti, B. Lepri, and M. Zancanaro. Multimodal recognition of personality traits in social interactions. In *Proc. Int. Conf. on Multimodal Interactions*, pages 53–60, New York, New York, USA, 2008. ACM Press.

[26] J. Shotton and A. Fitzgibbon. Real-time human pose recognition in parts from single depth images. In *CVPR*, 2011.

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

[28] A. Winters. Perceptions of Body Posture and Emotion: A Question of Methodology. *The New School Psychology Bulletin*, 3(2):35–45, 2005.