Design and evaluation of postural interactions between users and a listening virtual agent during a simulated job interview

David Antonio Gómez Jáuregui, Tom Giraud, Brice Isableu, Jean-Claude
Martin

Design and Evaluation of Postural Interactions between Users and a Listening Virtual Agent during a Simulated Job Interview

David Antonio Gómez Jáuregui*[1] | Tom Giraud[2] | Brice Isableu[3] | Jean-Claude Martin[4]

[1]Univ. Bordeaux, ESTIA INSTITUTE OF TECHNOLOGY, Bidart, France
[2]Human-Centered Multimedia, University of Augsburg, Augsburg, Germany
[3]Aix-Marseille Univ, PSYCLE, Aix-en-Provence, France
[4]Université Paris-Saclay, CNRS, LIMSI, Orsay, France

Correspondence
*David Antonio Gómez Jáuregui, Univ. Bordeaux, ESTIA INSTITUTE OF TECHNOLOGY, Bidart, France. Email: d.gomez@estia.fr

Summary

Postural interaction is of major importance during job interviews. While several prototypes enable users to rehearse for public speaking tasks and job interviews, few of these prototypes support subtle bodily interactions between the user and a virtual agent playing the role of an interviewer. The design of our system is informed by a multimodal corpus that was previously collected. In this paper, we explain how we were inspired by these video recordings of human interviewers to build a library of motion-captured movements that interviewers are most likely to display. We designed a fully automatic interactive virtual agent able to display these movements in response to the bodily movements of the user. Thirty-two participants presented themselves to this virtual agent during a simulated job interview. We focused on the self-presentation task of the job interview, the virtual agent being listening. Participants stood on a force platform that recorded the displacements of their center of pressure to assess the postural impact of our design. We also collected video recordings of their movements and computed the contraction index (CI) and the quantity of motion (QoM) of their bodies. We explain the different hypotheses that we made concerning (1) the comparison between the performance of participants with human interviewers and the performance of participants with virtual interviewers, (2) the comparison between mirror and random postural behaviors displayed by a female vs. a male virtual interviewer, and (3) the correlation between the participants' performance and their personality traits. Our results suggest that users perceive the simulated self-presentation task with the virtual interviewer as threatening and as difficult as the presentation task with the human interviewers. Furthermore, when users interact with a virtual interviewer that mirrors their postures, these users perceive the interviewer as being affiliative. Finally, a correlation analysis showed that personality traits had a significant relation to the postural behaviors and performance of the users during their presentation.

KEYWORDS:
virtual human, job interview, postural interaction, mirroring

1 | INTRODUCTION

When we speak about ourselves in front of others, we might perceive others' judgments as possible threats. The risk of failing a job interview induces anxiety and stress-related states (i.e., negative affects)[1]. Public speaking tasks and job interview simulations are relevant scenarios for training individuals to regulate their emotions. As a key step in everyone's life, job interviews are particularly stressful and engage our entire body with a strong arousal.

The design of virtual applications replicating these highly emotional situations would help individuals better manage their emotions and become less affected by these potentially traumatic moments. Accordingly, several researchers in human-computer interaction have proposed various models, prototypes and experimental protocols in an attempt to reach a sufficiently realistic level in terms of appearance, animation and interactive behaviors for public speaking and job interview training[2, 3]. Few of them consider the full body behaviors of users and focus instead on their speech, facial expressions, and physiology or the movements of body parts, such as hand gestures, head gestures and upper body movements. However, it is known that the full body transmits important affective signals[4] and is involved in strong interpersonal features, such as synchrony[5].

In this article, we explain how we used a multimodal corpus to inform the design of a listening virtual interviewer that is able to select its own body movements as a function of the user's full body movements. We describe a user study conducted with 32 participants simulating a self-presentation task during a job interview. In the proposed scenario, each participant must present him/herself to the virtual listener/interviewer in order to apply for a job. During the self-presentation, the virtual agent is able to produce postural behavior according to the postural behavior of the participant. In order to define the behavior of the listening virtual interviewer, we analyzed a video corpus of real job interviews that we collected previously. Based on this analysis, a motion corpus of postural moves was collected. Then, a system architecture is implemented in which a decision module decides the listening virtual interviewer's behaviors as a function of the user's postural behaviors captured from a force plate. We conducted a user study to assess the impact of the proposed postural interaction on users' behaviors and users' perception (e.g. perceived difficulty and perceived stress) during the self-presentation task. Our goal is to evaluate how different postural behaviors of the virtual interviewer can impact the performance of the participant during this task. For this study, two factors with two conditions were evaluated: 1) mirror vs. random postural behavior, and 2) female vs. male virtual agent. The postural behavior of the virtual interviewer is triggered by a change in the postural behavior of the user. "Random postural behavior" means that the virtual interviewer will reproduce a postural behavior different from the user's postural behavior. "Mirror postural behavior" means that the virtual interviewer will mirror the user's postural behavior.

The aim of the current study is thus to evaluate the relevance and feasibility of simulating postural interaction during a job interview. By simulating this postural interaction with virtual interviewers, we can determine if users are able to perceive the simulated self-presentation task as threatening and as difficult as the same self-presentation task conducted with human interviewers. In addition, we are willing to evaluate how the performance of participants can be impacted by different postural behaviors displayed by the virtual interviewer. The results of the current study may provide new guidelines to develop more realistic job interview simulators displaying personalized scenarios taking into account individuals' cognitive profiles in which the postural interaction is considered and modulated to increase or decrease the difficulty or perceived stress of the self-presentation task during job interview training.

The main contributions of this work are the following:

- A system architecture for managing postural interaction in a job interviewer scenario with a virtual interviewer.

- An evaluation of the system compared to a task with real interviewers showing its relevance for job interview training: the interaction is less stressful but perceived as threatening and difficult as a real task.

- A comparison of four different conditions of the system (female versus male, random versus mirroring postural behaviors) showing the specific effect of a mirroring female agent perceived as warmer.

- An exploratory analysis about personality traits.

This paper is organized as follows. Section 2 surveys related work. Section 3 explains how we designed and implemented our listening virtual interviewer. Section 4 describes a user study with 32 participants. The results are presented in Section 5 and discussed in Section 6. Section 7 concludes the paper and proposes future directions.

2 | RELATED WORK

2.1 | Psychological stress and performance assessment in public speaking situations

Lazarus defines psychological stress as the result of the comparison between the power of the environmental demands (stressor or load) and the psychological resources of the person to manage those demands[6]. Stress is therefore mediated by an appraisal process. Three types of psychological stress are identified: harm/loss, threat and challenge, which all result from different situation appraisals. Psychological stress and emotions are interdependent yet not often studied together. Multiple emotions can arise from stressful situations, for example: anxiety, fright, guilt, shame, frustration and sadness[7]. Stressful situations encompass a wide variety of contexts and may elicit a panel of negative emotions but also may possibly have some positive affects related to the challenging situation (e.g., excitement).

Several studies use a public speaking task to elicit stress. In this case, job interviews are quite relevant for eliciting social stress because the participant has to present himself and talk about personal matters[8]. The Trier Social Stress Test (TSST) was introduced in 1993[9] for the induction of moderate psychological stress in a laboratory setting for psychological and physiological studies. The TSST test consists of a simulation of a job interview: the participant takes ten minutes to prepare for a free five-minute presentation/speech and then delivers it in front of a jury. This protocol has been shown to induce stress impacting physiology, behaviors and cognition, although there are large individual differences[1]. Some authors have investigated the role of gender in influencing public speaking anxiety. Mejias et al.[10] found that female speakers are more prone to speaking anxiety than males. Similarly, Ralph et al.[11] reported higher anxiety patterns by female speakers in a gender-based study. In contrast, several authors[12, 13] found that gender does not have a significant effect on speaking anxiety. Regarding the gender of the interviewer, two studies[14, 15] found that the interviewer's gender can influence responses of the interviewee to some sensitive questions (e.g., feminism).

Psychological studies have shown that stress impacts expressions in speech[16], posture[17], facial expressions[18], physiology[9], postural control[19] and self-report[20]. Recent developments in multimodal data capture techniques have made it possible to detect these behavioral expressions during public speaking situations to assess the performance of participants. Ramanarayanan et al.[21] extracted several nonverbal behavioral cues (head pose, gaze, and facial expressions) and speaking proficiency features (prosody, pronunciation and grammar) from 17 participants during five public speaking tasks. The authors correlated these extracted features with human-rated scores to find the modalities that are useful in predicting public speaking performance. Aigrain et al.[22] proposed a system to perform automatic perceived stress detection by combining body postural behavior, facial features (Action Units) and crowd-sourced annotation. In a more recent work[23], they added a combination of physiological features (blood volume pulse, heart rate and heart rate variability) to improve the performance of their proposed system.

Recently, some authors have used automatic detection of non-verbal visual or audio behavioral cues to help speakers to be conscious of their behavior during job interviews. Nguyen et at.[24] proposed a method to predict hiring decisions using non-verbal visual and audio behavioral cues extracted from thin slice interactions in a collected dataset of 62 real job interviews. Tanveer et al.[25] proposed an interface to automatically extract the repetitive patterns of human gestures to make speakers aware of their mannerisms. Naim et al.[26] proposed a computation framework for automatically quantifying verbal and nonverbal behaviors in the context of job interviews. The extracted behaviors are used to predict the ratings for interview traits, such as excitement, friendliness, and engagement.

All these studies have shown that extracting multi-modal cues (voice, posture, gestures, facial expressions, etc.) during public speaking tasks such as job interviews, is relevant to assess psychological stress and therefore, to develop systems that help users to improve their performance under stressful situations.

2.2 | Virtual agents for public speaking and job interview scenarios

Simulation of virtual audiences for public speaking scenarios has received a large amount of interest from the research communities in Affective Computing, Social Computing and Virtual Reality. Virtual audiences were used to help musicians practice, to reduce social phobia with exposure therapy, and more generally, to study the effect of stress in demanding situations[27]. The use of a virtual environment populated by a human-like audience to provoke anxiety and treat social phobia is a frequently studied research case[28]. Often having poor-fidelity appearances, animated virtual audiences have. been observed to have positive long-term effects, with behavioral animations being a key determinant of the success of the therapy. The main advantage of these virtual audiences over real-world workshops is the possibility to provoke fear and anxiety in a controlled manner. By slightly

modifying the audience behaviors and attitudes, researchers can generate a gradual exposure[27]. Several researchers have investigated the impact of neutral, negative or positive feedback from the audience on users' performances[29, 30], including even bored or critical attitudes displayed by the virtual attendees[27]. Most often, the behaviors displayed by the audience are controlled by an experimenter (Wizard-Of-oZ) or are pre-scripted but not computed in real-time as a function of the user's performance or the user's nonverbal behaviors.

Recent research projects go further by developing interactive characters in scenarios such as job interviews. MACH is a conversational coach that provides graphical and textual feedback to participants on the overall job interview performance based on automatic analysis of users' facial and voice cues[3]. The interaction is a sequential dialog with some mirroring and random behaviors displayed by the virtual agent. An evaluation of the system showed that users who received such feedback regarding their interaction improved their performance thanks to this training. The TARDIS European project proposes a serious interactive game platform where the participant interacts with a virtual interviewer: the system infers the user's emotions and states of mind based on various nonverbal cues (face, voice, gestures and postures) and then generates nonverbal behaviors through an affective module following a sequential dialog scenario[2]. Some aspects of the project concentrate on the reasoning process (Theory of Mind)[31], while others explore different styles of virtual interviewers. A positive interviewer can display signs suggesting that he understands and supports the user to augment the user's self-efficacy and create a warm and friendly climate. A demanding recruiter can be used instead to train users to manage more difficult situations and improve their coping abilities (although the project itself is not focused on social stress)[32]. Another interview scenario is investigated in[33], with the goal being to make the interlocutor feel comfortable to elicit self-disclosure to detect depression and anxiety disorders. More recent studies[34, 35] proposed an interactive virtual audience for public speaking training. The virtual audience's behavior is driven by multimodal signals extracted from users, such as the amount of gaze directed at the audience, body activity, voice quality and heart rate. A recent study[36] investigated trust and acceptance of outpatients interacting with a virtual psychiatric agent. The virtual agent was perceived as trustworthy and well accepted by the patients, confirming the good engagement of patients in the interaction.

When candidates present themselves, interviewers display listening behaviors. Virtual humans as listeners are an active domain of research. Backchannels are modeled following data-driven approaches (machine learning)[37] or theory-driven approaches (e.g., rule-based models)[38]. In machine learning approaches, two of the most popular methods are Hidden Markov Models (HMM) and Conditional Random Fields (CRF)[39], sequential probabilistic models providing real-time probabilities of the apparition of discrete nonverbal signals. Models are trained on annotated datasets using various nonverbal signals (voice, head movements, and facial expressions)[37]. With the development and efficacy of data-driven approaches, the tendency is to propose multimodal models[37] with efforts toward the collection of more behaviors[40]. Some models and prototypes of virtual listeners focus on the reactive aspect of backchannels and propose models of generic responses[39]. In these cases, no understanding of the narrative (e.g., the user's speech) is required. Another approach considers higher-level processes, such as state of mind and affects, to guide backchannel behaviors. In these cases, specific responses tied to the narrative structure are considered. Wang et al.[41] proposed utilizing varied goals and roles to go "deeper" than just generic feedback. Buschmeier et al.[42] attempted to determine the timing of feedback by considering "information need" of the speaker, which necessitates keeping track of the narrative. In the SEMAINE project[43], recognition modules of emotional state and the user's words are added to guide the generation of specific responses.

Overall, these projects tend to integrate voice, face, and gestures to infer emotions and intentions, but they still do not consider the postural control of the user's full body in their models. In addition, none of the projects simulating virtual listeners integrate postural movements as an interactive listening modality.

2.3 | Postural control in stressful and interpersonal situations

In his Posture Scoring System, Bull defines a posture as "any movement which is taken up and maintained for at least one second"[44].These static body configurations can convey specific emotions[45], action tendencies[46] or or the intensities of some emotions[4]. Unlike the traditional static view of posture, postural control is dynamical. In motor control studies, posture is defined as "the body automatic stabilization in the gravity force field in a standard position specific to the species"[47]. Postural control, often synonymous with balance, is therefore the dynamical process by which equilibrium is obtained by maintaining the vertical projection of the body's center of gravity within the base of support (the area under the feet). Our body is always swaying around its equilibrium point, and this sway can be impacted by different variables, such as aging or attention[48]. The most common way to quantify human body sway is by extraction of the Center Of Pressure (hereafter called COP) trajectory through a force

platform that measures the ground reaction forces. The COP is therefore "the position of the resultant vertical component of the ground reaction force applied to the body at the ground surface"[48].

Affects' impact on postural control is a relatively new field of research in the motor control domain[49]. Hypotheses regarding the impact of emotions on body signals are often driven by motivational theories[50] or the action tendency component of appraisal theories[51]. Negative or threat stimuli can trigger flee behaviors, fight behaviors or freezing behavior[52]. These approach and avoidance reactions are present in quiet standing as well as when initiating a walk (stepping forward or stepping back)[53].

Posture mirroring and synchrony within conversations between strangers is not a recent research domain[54, 55]. Postural mirroring is defined as the mimicking of another individual's stance, gestures, or other nonverbal behaviors, whether fully or partially[56]. Several studies have found that when a stranger mimics or mirrors the postural behaviors of somebody during a social interaction, higher positive judgment and greater affiliation of the mimicker is found[57, 58]. In addition, it has been shown that mirroring behavior facilitates the smoothness of interactions and increases liking between interaction partners[59]. More recently,[60] used modern methods of quantification of postural dynamics to measure postural coordination. The rationale behind the idea that the body sway of both participants might be coordinated is that postural stabilization is not an "end goal process" but a "facilitator of the goal of supra-postural activity." The center of pressure has to constantly adjust to required center of mass changes. With conversation considered as a supra-postural task that involves listening, gesturing and visual inspection, they investigated the coordination pattern within a puzzle task involving conversation. In a subsequent study, these authors replicated their findings and described interpersonal coordination as a set of emerging structures[61]. Postural control appears to be a dynamic and complex process tightly related to emotional experiences and social interactions, in accordance with the theory that emotions are embodied in our tendencies to act and stand in the environment. As such, and although it has not been considered in past virtual listeners, it appears to be a privileged channel to be analyzed in such a highly emotional interactive situation.

Accordingly, we aim to endow a virtual listening interviewer with postural behaviors. We also propose the implementation of mirroring mechanisms between the participant's and the listening virtual interviewer's postural moves to reproduce spontaneous coupling behaviors. These implementations are evaluated in a user study having the same configuration as the real-life corpus that we collected[62]. The job interview simulator is evaluated both against reality and between its configuration variables (mirror versus random behaviors and female versus male virtual interviewers).

3 | SYSTEM DESIGN

The job interview simulator presented in this paper aims to simulate the presentation phase of an interview: a candidate presents him/herself standing in front of a seated virtual interviewer who is listening. The second choice of this research is to focus on whole-body postural moves in the interaction (standing for participants and seated for virtual interviewers). Analysis of a video corpus of job interviews that we collected previously helped us define these behaviors. The postural behavior of the listening virtual interviewer was defined by visually inspecting and annotating manually the postural behavior of the job interviewers. In this section, we first briefly describe this job interview corpus. Then, we detail the design of the virtual interview simulator and present the three modules: rendering, sensing and decision.

3.1 | Job Interview Corpus

The corpus of simulated job interviews with real humans is a multimodal database[62]. 19 Participants were subjected to a public speaking task adapted from the Trier Social Stress Test[9]. These 19 participants were recruited from University of Paris-Sud (male n=7, 37%; female n=12, 63%). The average age of participants was 26 years (SD= 6.1). All participants were volunteers and signed an informed consent designed in collaboration with the administrative heads of the partners' laboratories. Participants were told that they had to present themselves in front of two human interviewers (a man and a woman interviewer). The instructions provided to participants were as follows: "You are applying for a job that is particularly important to you. You will have five minutes to present yourself in front of a camera and describe your personal characteristics (e.g. personality, skills, and experiences) that you view as strengths and that would allow you to be relevant for this job. You must also indicate those characteristics that might be perceived as weaknesses and how you did or are willing to improve them. Be as sincere as possible. This is an important prerequisite for this position."

Participants were standing in front of the two seated interviewers. The task consisted of five phases: a two-minute reading of a one page text (reading phase), a one-minute instruction reminder (listening phase), a five-minute self-introduction speech

(presentation phase) and two feedback sessions from the two interviewers (dialogue phase). The assessors' feedback and questions were regarding the participant's performance. The first interviewer provided negative feedback. The second interviewer provided positive feedback. Each participant received the negative feedback first and the positive feedback second (to have the participants end the experiment in a positive state). This protocol is a modified version of the widely used Trier Social Stress Test (the original protocol is briefly described in section 2.1) known to induce moderate stress in laboratory settings. This TSST protocol was modified so as to induce stress variability. The main modification was to add a feedback phase between the judges and the participant, one judge providing positive feedbacks, and the other providing negative feebdacks. In this paper, we do not consider this dialogue phase and focus on the self-presentation phase which is in line with the TSST protocol. Behaviors of 32 participants were recorded via a multimodal setup including speech, video of their facial expressions and body movements, balance via a force platform, and physiological measures. Questionnaires were used to assert emotional states, personality profiles and relevant coping behaviors to study how participants cope with stressful situations. Several subjective and objective measures were also collected. Specifically, relevant to this paper for comparison purposes, threat and challenge (ThCh) appraisals were assessed before the task via four items . Threat and challenge (ThCh) appraisals were measured just before the task by 4 questions on a 7-point Likert scale[63]. Two items assess the primary appraisal which is how important and stressful participants perceive the event (Threat). The other two items assess the secondary appraisal which is how well participants think they will cope with the task (Challenge). A ratio of the two scores (primary/secondary) generate appraisal scores, where higher scores denote threat (Apprai R). In addition, perceived difficulty and stress were assessed after the task[20].

To limit the range of interpersonal interactive phenomena to be analyzed, we focus on the presentation phase of the job interview. During these five minutes, the participant presents him/herself to the interviewers and explains the reasons for his/her application. Modeling this phase of the interview enables us to avoid considering dialog-related phenomena and focus on one type of asymmetrical interaction: the participant talks (standing) and the interviewers listen (seated). As in most job interview situations, the interviewers in our corpus were seated in front of the participant. As we are focused on the presentation phase, the interviewers are only listening to the participant without any verbal interventions. A first inspection of the interviewers' videos revealed that the behavioral stream of bodily movements (face excluded) may be better characterized by a series of discrete behaviors than a continuous chain of movements. Although we are in a seated configuration, this analysis is in line with Duarte et al.[48] who characterized unconstrained standing by specific sequential patterns (i.e. fidgeting, shifting, drifting) rather than continuous movements. Thus, we performed an automatic segmentation of behaviors to extract all these sequential interviewers' behaviors from the interviewers' video sequences. This automatic segmentation was performed using the motion history images algorithm (MHI)[64], which provides an image template that shows the recency of motion for each frame in a sequence. Thus, for each image of the video sequence we extracted the number of pixels in which motion was detected (quantity of motion). Then, for each video sequence, our automatic segmentation algorithm extracted only the segments of the video sequence in which the number of pixels is larger than a threshold (chosen experimentally from visual inspection). Once the segments were automatically extracted, we performed a manual inspection of these video segments in order to remove "irrelevant" segments (e.g., movement due to the disturbance of an external source). 260 segments corresponding to postural moves remained, mainly corresponding to shift in position or fidgeting (i.e. moving while remaining in the same position). This extraction procedure also provides us some elements about the movement's characteristics used to design the listening virtual interviewer: each postural transition move lasts for approximately 5 seconds. with a postural move appearing approximately every 30 seconds.

3.2 | Job Interview Simulator

The system architecture is composed of three modules: the rendering module that displays and animates the listening virtual interviewer in a virtual interview environment, the sensing module that captures in real-time the user's postural behaviors, and a decision module that links the sensing and the rendering modules to decide what and when to play the listening virtual interviewer's behaviors as a function of the user's behaviors (Figure 1). Each module is explained in the following subsections.

3.2.1 | Rendering module

The male or female virtual interviewer is seated and taking notes on a laptop during the interview (a typical configuration in meetings today). The face of the listening virtual interviewer is hidden using a blur filter that follows its head movements. The purpose of this filter is to keep the properties of the face visible enough (location of the eyes, mouth, etc.) but reduce these pieces of information to disable the possibility of discerning eventual facial movements or incongruent facial expressions since our focus is on the interviewer's postural moves. The decision to blur the face is supported by neuroscience and psychological
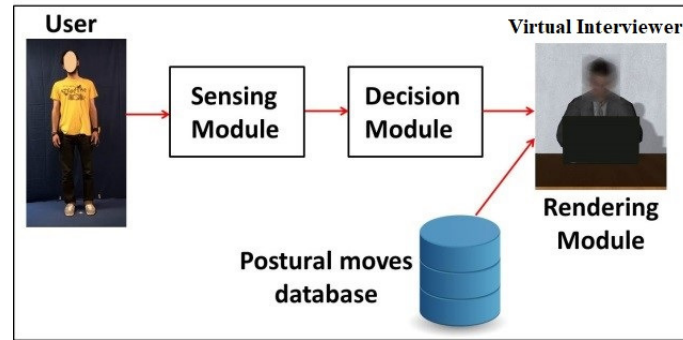
FIGURE 1 System architecture for managing postural interaction during the job interview simulation.

studies showing that bodily expressions are not easily recognized when facial expressions are not congruent with bodily behavior[65, 66]. In addition, several studies have investigated the effect of face blurring during body language processing. These studies have shown that body language processing is performed automatically and unconsciously and does not rely uniquely on facial expressions[66–68].

Postural moves of the virtual interviewer need to be varied to give him/her some credibility. Based on the detailed inspection of the video extracted from the corpus, we defined possible postural moves in the two-dimensional horizontal plane that depend on the position of the listening virtual interviewer. We defined 7 possible whole-body *postural positions* (Center, Front, Front-Left, Front-Right, Back, Back-Left, and Back-Right) that the listening virtual interviewer can take. To further deepen the variability of the virtual interviewer, we defined two types of *postural moves* that correspond to significant moves of interviewers that we observed in our job interview corpus:

1. Transitions: Anterior-Posterior (AP) and Medio-Lateral (ML) moves from one position to another.

2. Fidgets: moves in the same position.

In addition to fidget moves and transition moves, we included some idle moves. This means that no significant movement is played but movement still occurs in a subtle way so as not to appear totally static. An idle movement is assigned for each position.

All possible moves of the virtual interviewers were motion captured by a non-professional male actor to ensure realistic seated postural moves (Figure 2). In each of the seven positions, we captured three different fidgets. For each possible transition, we captured two moves in one direction (position A to position B) and two other moves in the other direction (position B to position A). All these moves have duration ranging from three to eight seconds to match observations from the corpus. This provided the virtual interviewer animation module with 69 possible postural moves. As, on average, an interviewer performs about eight significant postural moves within a presentation, this provides enough variability to avoid having the user perceives that two moves are the same. In addition, for each postural position, we captured two takes of 40 seconds of so-called idle movement (i.e., doing nothing). Finally, two motion sequences were captured to start and end the presentation task. In the starting motion sequence, the interviewer provides the instructions of the presentation task to the user. In the ending motion sequence, the interviewer thanks the user for his/her participation. Figure 3 shows the seven possible postural positions and transitions obtained from the motion capture session.

To provide smooth realistic animation between each interviewer postural behavior (fidget or transition), the rendering module makes use of a state machine in which each animation transition is interpolated by generating intermediate animations at runtime (using the Mecanim Animation System from Unity3D[69]. Two virtual humans were animated: one businesswoman and one businessman. The seven postural positions performed by our virtual businessman are displayed in Figure 4.

3.2.2 | Sensing module

The sensing module collects signals from different devices and computes several user behavioral measures in real-time. We use the Social Signal Interpretation (SSI) framework, which supports streaming from multiple sensors and includes mechanisms for their synchronization[70].

The user's postural behavior is collected from a force platform device on which the user stands. The force platform device (Amti AccuGait) is capable of providing, in real-time (60 Hz), the three ground reaction force components (Fx, Fy, Fz) and

FIGURE 2 Recording the postural moves for the virtual interviewer from a motion capture session.
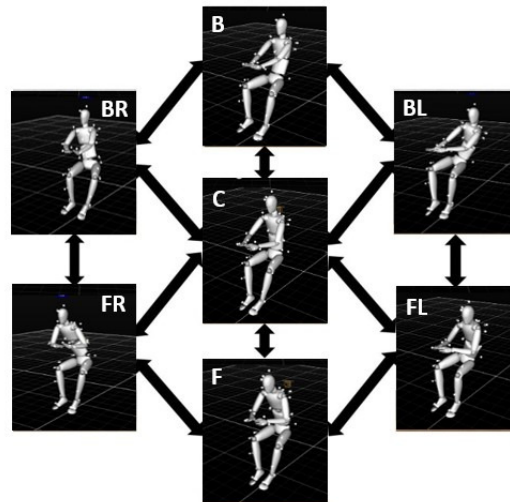

FIGURE 3 Postural positions and transitions recorded for the virtual interviewer. Each postural position is shown with its respective nomenclature (C = Center, BL = Back-Left, B = Back, BR = Back-Right, FL = Front-Left, F = Front, and FR = Front-Right).

the three moment components (Mx, My, Mz) that are measured as a user is standing and is in contact with the platform. These measures are used to compute the Center Of Pressure (COP) displacements along the two axes, Anterior-Posterior (AP) and Medio-Lateral (ML), using the following equations:

$$COP_x = -\frac{M_y}{F_z}.$$
(1)

$$COP_y = \frac{M_x}{F_z}.$$
(2)

These measures (COP_x and COP_y) provide information about the user's ML and AP postural behavior (postural control) while standing on the force platform. In our system, both measures are sent in real-time to the decision module using a socket connection.
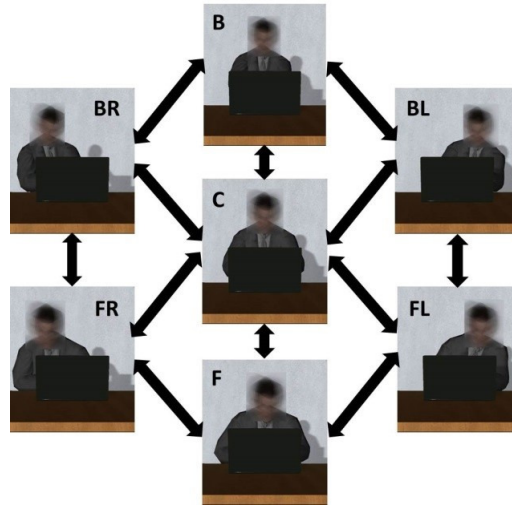
FIGURE 4 Postural positions of the virtual interviewer (C = Center, BL = Back-Left, B= Back, BR = Back-Right, FL = Front-Left, F = Front, and FR = Front-Right).

3.2.3 | Decision module

This module couples the sensing and rendering modules by implementing a rules-based matching algorithm that produces the postural behavior of the virtual interviewer from the postural behavior of the user (how postural moves of the virtual interviewer are influenced by the user's postural moves). This module receives as input the horizontal and vertical displacements of the user's center of pressure (COP_x and COP_y) and decides the postural behavior that the virtual interviewer will perform. The interviewer postural behavior can be either mirror (the same postural move as the user) or random (random selection of the postural move).

To select the appropriate behavior of the virtual interviewer, the module must first identify the type of postural behavior displayed by the user. As explained previously (section 3.2.1), two postural behaviors are considered when analyzing the user's moves: fidget and transition. Each type of postural behavior is detected by comparing the displacements of the user's center of pressure (COP_x and COP_y) between a current time t and an initial time t_0.

A *fidget behavior* is detected if the displacements of COPx and COPy meet the following three criteria:

1. $if(\sum_0^t COP_x + \sum_0^t COP_y) > T_{acc}$

2. $if |COP_{xt} - COP_{x0}| < T_f$

3. $if |COP_{yt} - COP_{y0}| < T_f$

$\sum_0^t COP_x$ and $\sum_0^t COP_y$ represent the accumulated displacements of COP_x and COP_y between the current time t and the initial time t_0. COP_{xt} and COP_{yt} are the horizontal and vertical COP displacements at the current time. COP_{x0} and COP_{y0} are the horizontal and vertical COP displacements at the initial time. T_{acc} and T_f are thresholds, these values were empirically set during pretests (about 5 persons tested the system). These values were obtained by computing the average displacement (COP_x and COP_y) from several fidget behaviors of a user standing on the platform and presenting himself to the virtual interviewer.

A *transition behavior* is detected when the user's postural position at the current time t is different from the user's postural position at the initial time t_0. To identify the user's postural position, we computed the orientation of the user's center of pressure in the XY plane using the following conditions:

- $X = 0, \ Y = 0$
- $if(COP_x > T_{disp}) \ Then \ X = +$
- $if(COP_x < -T_{disp}) \ Then \ X = -$
- $if(COP_y > T_{disp}) \ Then \ Y = +$
- $if(COP_y < -T_{disp}) \ Then \ Y = -$

TABLE 1 Postural positions associated with the user's horizontal (X) and vertical (Y) orientation.

X	Y	Postural Position
0	0	CENTER
0	+	FRONT
+	+	FRONT-RIGHT
-	+	FRONT-LEFT
0	-	BACK
+	-	BACK-RIGHT
-	-	BACK-LEFT

TABLE 2 Postural positions associated with the user's horizontal orientation and COP vertical displacement.

X	Y	COP_y	Postural Position
+	0	$COP_y>0$	FRONT-RIGHT
-	0	$COP_y>0$	FRONT-LEFT
+	0	$COP_y<0$	BACK-RIGHT
-	0	$COP_y<0$	BACK-LEFT

T_{disp} is a threshold chosen during pretests (about 5 persons tested the system). These values were obtained by computing the average displacement (COP_x and COP_y) from several transition behaviors of a user standing on the platform and presenting himself to the virtual interviewer. Once the orientation is found, the postural position of the user is identified using the following two tables (Table 1 and Table 2):

To support postural interaction (mirror or random) with the virtual interviewer behavior (rendering module), the decision module uses the information about the user's current postural behavior (sensing module) to detect a *fidget behavior* or a *transition behavior*. Once the behavior was detected, the decision module sends a message to the rendering module each time that a virtual interviewer behavior or idle movement must be played. The idle movement is played in the case of a *fidget behavior* or a *transition behavior* is not detected.

In a last step, we checked if the distribution of the produced postural moves by our algorithm in the case of our 5 beta-testers corresponds to the one present in the real corpus. With a postural move every 20 seconds, the algorithm produce more postural moves than in the real corpus (one every 30 seconds). This difference might have been caused by the fact that our beta-testers were aware of the purpose of the setup and thus more inclined to perform postural moves.

4 | USER STUDY

We conducted a study to assess the impact of the proposed postural interaction on the users' behaviors and users' perception (e.g. perceived difficulty or perceived stress) during the self-presentation task. In addition, as stated in section 2.1 and 2.3, several studies have shown that interviewers' gender [14, 15] and mirroring behavior [57–59] can influence performance and perception of participants during public speaking situations. Thus, the impacts of these two variables are also evaluated in our study. Specifically, this study aims to answer the following questions:

- Is the self-presentation task with virtual interviewers as threatening as the self-presentation task with human interviewers?

- How is the performance (postural behavior and perception) of participants affected by the postural behavior and the gender of the virtual interviewer?

- Is there a relationship between the performances of participants during the self-presentation simulation and their personality traits?

Participants were asked to stand on a force plate and present themselves in front of the virtual interviewer to apply for a job. Subjects did not know that their balance was detected by the force plate. To animate the virtual interviewer, we implemented
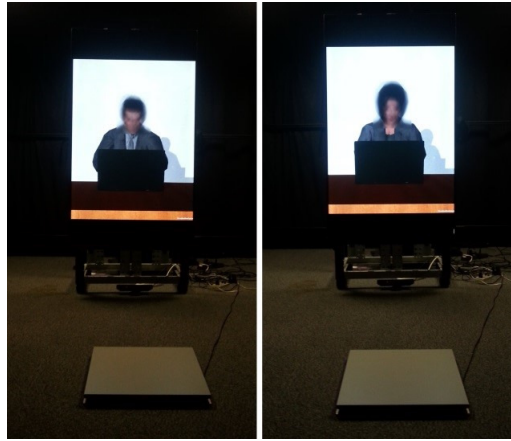
FIGURE 5 Setup and virtual interviewers displayed during the presentation task. The left image shows the male virtual interviewer. The right image shows the female virtual interviewer.

four conditions resulting from the combination of two variables: 1) mirror vs. random postural behavior, and 2) female vs. male virtual agent. As mentioned earlier (introduction, section 2.3 and section 3.2.3), the postural behaviors evaluated in our study are triggered by the change of postural behavior of the user. However, mirroring behavior refers only to the virtual interviewer mirroring the same postural move (like in a mirror) of the participant[56].

4.1 | Apparatus and Participants

An AMTI Accugait force plate was placed one meter in front of a 75" screen (Figure 5). Since we are interested in measuring the participants' postural behavior during the self-presentation task, a Microsoft Kinect was placed under the screen to collect measurements of these postural behaviors (these behavioral measures will be explained in section 4.4.1). In our setup, the force plate and Kinect data were used to record users' postural behavior. The force plate data was also used by the decision module to decide the postural behavior of the virtual interviewer. The virtual scenario was implemented using Unity3D and displayed with a resolution of 1920 x 1080 pixels at a frame rate of 60 FPS. A male virtual interviewer and a female virtual interviewer were used (Figure 5).

Thirty-two participants aged from 21 to 51 (x = 29.95, SD = 6.45), sixteen males and sixteen females, participated in the experiment. 25 participants were French, 4 participants were Algerians, 2 participants were Tunisians and 1 participant was Lebanese. All participants were fluent in French (the instructions and all questionnaires were written in French). Each participant signed a consent form which included information about the study and explained that data was collected and held anonymously.

4.2 | Procedure

Participants were instructed to present themselves to a virtual interviewer in order to apply for a job that was particularly important for them. These first and short instructions (given by the experimenter) were given with the aim of briefly informing the participant about the main task of the experiment. Before the task, participants were asked to answer a personality questionnaire (Big Five[71]) and a questionnaire assessing threat and challenge (ThCh) appraisals[63]. Then, the participant entered the experimental room and was asked to stand in the middle of the force plate in front of the virtual interviewer (Figure 7). The experimental room was sound isolated, and a low level of illumination was kept constant throughout the experimental session. Once the participant was ready and the instructions were clear to them, the experimenter left the room and initialized the virtual interviewer behavior remotely. Immediately, the virtual interviewer provided the participant with the following detailed instructions (in french language) for the presentation task: "You are applying to a job that is particularly important to you. You will have five minutes to present yourself in front of a camera and describe your personal characteristics, for example, your personality, skills, and experiences, that you view as strengths and that would enable you to be relevant for this job. You must also indicate those characteristics that might be perceived as weaknesses and how you did or are willing to improve them. Be as sincere as

Time	Human interviewer study	Virtual interviewer study
Before the task	• Personality questionnaire (Big Five) • Appraisal (ThCh) questionnaire • Instructions from the experimenter	• Personality questionnaire (Big Five) • Appraisal (ThCh) questionnaire • Instructions from the experimenter
Public speaking task	• Reading of a 1 page text (*reading phase*) • Instruction reminder (*listening phase*) • 5 minutes self-introduction speech (*presentation phase*) • 5 minutes questions and feedback from two interviewers (*dialogue phase*)	• Instruction reminder (*listening phase*) • 5 minutes self-introduction speech (*presentation phase*)
After the task (short questionnaires)	• Perceived difficulty and stress • Emotional response	• Perceived difficulty and stress • Emotional response • Perception of the virtual interviewer (2 questionnaires) • Movement perception

FIGURE 6 Experimental steps of the human interviewer and virtual interviewer studies.
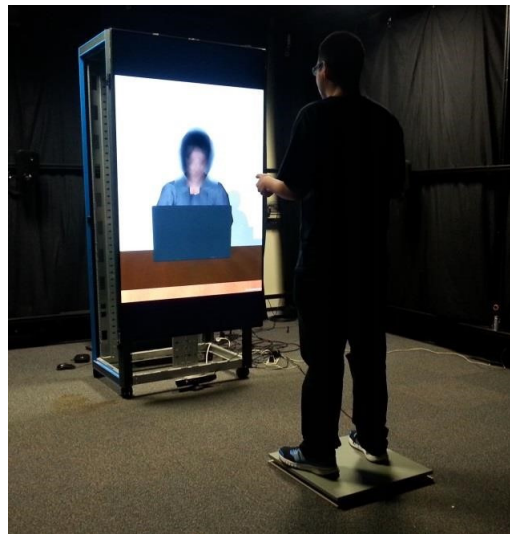
FIGURE 7 The setup designed for the presentation task. The participant speaks to a virtual interviewer to apply for a job. Behavioral measures are recorded from a Kinect camera and force plate sensors.

possible. This is an important prerequisite for this position." These instructions were the same as the instructions used for the initial corpus collection[62].

The presentation task lasted five minutes per participant. During the task, behavioral measures of the participant were extracted from the Kinect and the force plate sensors. If the participant finished his/her presentation before the end of the five minutes, he/she was instructed to wait for an alarm sound indicating the end of the presentation phase. At the end of the five-minute interval, the virtual interviewer thanked the participant and allowed the participant to leave the experiment room and join the experimenter. Finally, five short questionnaires (described in section 4.4.2) were given to the participant to evaluate his/her subjective perception of the virtual interviewer (humanness, dominance and warmth), perceived stress and difficulty, emotional response and movement perception during the presentation task. Each participant performed the experiment in one single session lasting approximately 20 minutes overall.

The figure 6 shows the experimental steps for virtual interviewer and human interview studies. As shown in this figure, the *reading phase* and the *dialogue phase* were not replicated in the virtual interviewer study. The participants of both studies answered the same questionnaires before and after the task. However the participants of the virtual interviewer study answered three more questionnaires related to the perception of the virtual interviewer (humanness, dominance and warmth) and movement perception.

4.3 | Experimental design

A between-subject experimental design was used. Two independent variables were used for the virtual interviewer: *Movement* and *Gender*. For the *Movement* variable, two conditions were considered: *mirror* and *random*. For the *mirror* condition, the virtual interviewer performed the postural change in the same direction (as in a mirror) as the participant. In the *random* condition, the virtual interviewer performed a random postural change. Both conditions were performed when a postural change of the participant was detected (section 3.2.2). For each experiment, participants were not informed about the postural changes of the virtual interviewer. For the *Gender* variable, two conditions were considered: *male virtual interviewer* and *female virtual interviewer*. Both virtual interviewers played the same animation. In our experimental design, the 32 participants were divided in four groups. Each group of participants was evaluated on each of the following combination of variables: 1) *mirror* and *male*, 2) *mirror* and *female*, 3) *random* and *male*, 4) *random* and *female*.

4.4 | Measures

Behavioral measures and self-reports were collected from the participant. Behavioral measures are related to the postural and body behavior of the participant during the presentation task. Participants are not informed that their behavioral measures are collected during the experiment. Self-reports are related to the feelings of the participant before and after the presentation task and his/her subjective perception of the virtual interviewer. All collected measures are described below.

4.4.1 | Behavioral measures

Four behavioral measures were collected during the presentation task: the contraction index (CI) and the quantity of motion (QoM) extracted from the Kinect, ML displacements of the center of pressure (COP_x) and AP displacements of the center of pressure (COP_y) extracted from the force plate. COP_x and COP_y are described in Section 3.2.2. CI and QoM measures are computed from RGB-D Kinect data, and both measures are described below.

The *Contraction Index* (CI) is a value, ranging from 0 to 1, indicating the degree of expansion / contraction of the body in the 2D plane of the camera. It is considered to be a behavioral indicator reflecting the emotional effusiveness of performers[72]. The larger the CI (close to 1), the more closed the stance of a performer; inversely, the smaller the CI (close to 0), the more open the stance of the performer. The CI is computed as the ratio of the area of the bounding box that covers the human silhouette and the area covered by this silhouette[72, 73]. In our system, the human silhouette is extracted from Kinect depth data using the player detection algorithm provided by the Kinect SDK[74].

The *Quantity of Motion* (QoM) is a value between 0 and 1 that estimates the total amount of motion of the body in the 2D plane of the camera. It is often considered a measure of kinetic energy and recurrently associated with emotional arousal in affective situations[75]. We used the motion history images algorithm (MHI)[64], which provides an image template that shows the recency of motion in a sequence. Both measures (CI and QoM) are computed using the following equations:

$$CI = \frac{A_s}{C_s}. \tag{3}$$

$$QoM = \frac{QoP}{C_s}. \tag{4}$$

where A_s is the area of the silhouette (in pixels), C_s is the area of the bounding box surrounding the silhouette, and QoP is the quantity of pixels moving in the image (figure 8.

4.4.2 | Self-reports

Two questionnaires were collected from the participants before the presentation task to identify dispositions influencing the performance in regulating stress during the simulation.

The Big Five Inventory (BFI) is the most widely used and extensively researched model of personality. It is a self-report inventory designed to measure five broad factors (dimensions) of personality traits: Openness to experience (O), Conscientiousness (C), Extraversion (E), Agreeableness (A), and Neuroticism (N)[76]. (John and Srivastava, 1999). For this study, the French version is used, which consists of a 45-item questionnaire scored on a 5-point Likert scale[71].

Threat and challenge (ThCh) appraisals were measured by four questions on a 7-point Likert scale[63]. Two items assess the primary appraisal, which is how important and stressful participants perceive the event to be (Threat). The other two items
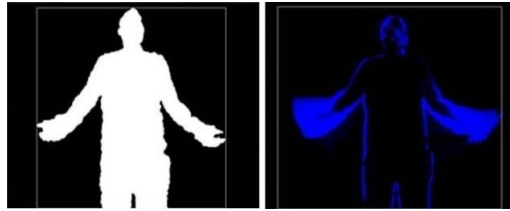
FIGURE 8 Behavioral measures. Left image: the contraction index (CI) computed from the human silhouette and its bounding box (white rectangle). Right image: the quantity of motion (QoM) computed from the motion history image (blue pixels) and the silhouette's bounding box.

assess the secondary appraisal, which is how well participants think they will cope with the task (Challenge). A ratio of the two scores (primary/secondary) generates appraisal scores. Higher scores denote a threat. This questionnaire had also been collected in the initial human-human interactions. The following three questionnaires regarding how participants perceived the virtual interviewer were collected after the presentation task:

1. *Movement perception* is measured by three items on a 5-point Likert scale (Annex 1). The first item measures the self-movement perception during the task. The second item measures how the participant perceived the movements of the virtual interviewer. The third item measures if the participant felt that the movements of the virtual interviewer were similar to his own. The items aimed to control if the mirroring behaviors were perceived: mirroring has to be unconscious to impact the interaction[77].

2. *Perceived humanness scores*[78] are motivated by the original graph of the uncanny valley proposed by[79]. These indices measure participants' attitudes and acceptance toward anthropomorphic characters. The six following items are used: artificial-natural, human made-humanlike, without definite lifespan-mortal, inanimate-living, mechanical movement-biological movement, and synthetic-real. The perceived humanness score is computed from the average of these six items. The reliability of this test was measured, obtaining a Cronbach's coefficient alpha of 0.83 (good internal consistency).

3. *Perceived dominance and warmth*[80] scores are used to measure social perception of human behavior. Two items were used: dominant-submissive and warm-cold.

The following two questionnaires regarding how participants felt during the interaction were collected after the presentation task:

1. *Perceived difficulty and stress*[20] assess the perception of participants of their own performance in a speaking task. In the current study, we selected two relevant questions to measure stress and difficulty on a 5-point Likert scale. This questionnaire was also collected in the initial human-human corpus.

2. The *Self-Assessment Manikin* (SAM)[81] is a non-verbal pictorial assessment technique used to measure each of the three major emotional dimensions (pleasure, arousal, and dominance) associated with the participant's emotional response after the presentation.

5 | ANALYSES AND RESULTS

In this section, three analyses are presented. In the first analysis, a comparison between the collected measures obtained with the virtual interviewers and the measures obtained with human interviewers is presented. In the second analysis, the performance of participants during the presentation task is examined with respect to the variables (*Male/Female* virtual interviewer and *Mirror/Random* behavior). Finally, a correlation analysis between the participants' performances and their personality traits is described.

5.1 | First Analysis: Virtual Interviewer vs. Human Interviewer
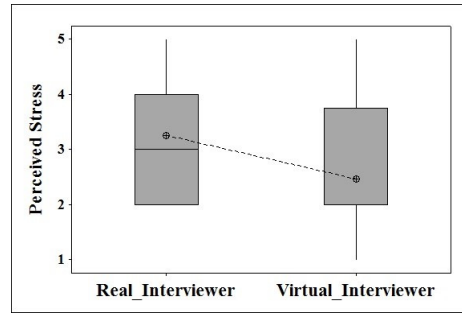
FIGURE 9 Box plot for perceived stress; dashed lines connect mean values. Participants reported higher perceived stress scores with the *Real Interviewer* condition.

We expected that human interviewers would induce in participants a higher perceived stress and difficulty during the self-presentation task than the virtual interviewers since the presence of real people induces the possibility of reasoning and online social judgments that are difficult to implement in a virtual interviewer. We considered the subjective and behavioral performance of participants during the presentation task for each condition (*Real interviewer* and *Virtual interviewer*).

Regarding the subjective performance, a one-factor between-subject ANOVA was conducted for each of the following subjective scores obtained from each participant: appraisal scores (ThCh), perceived difficulty scores and perceived stress scores. Regarding the behavioral performance, a one-factor between-subjects ANOVA was conducted for each behavioral measure considered (section 4.4.1). For each participant, the average Contraction Index (CI), the average Quantity of Motion (QoM), the standard deviation of horizontal and vertical displacements (COP_x and COP_y) were computed from the entire duration of the presentation task. For comparison purposes, the four behavioral measures from the virtual interviewer experiment were extracted using the same algorithms and with similar conditions (controlled lighting conditions and same camera angle) to those used in the job interview corpus with human interviewers (described in section 3.1). As shown in figure 8, Contraction Index (CI) and Quantity of Motion (QoM) measures were obtained using image processing algorithms[64, 74]. For each ANOVA analysis, the independent variable (factor) was the type of interviewer: *Real Interviewer* or *Virtual Interviewer*.

We expected that the participants' perceived threat before the self-presentation task with virtual interviewers, would be similar to the participants' perceived threat with real interviewers. However, during the self-presentation task, the real interviewers may induce a more difficult and stressful experience since, as mentioned earlier, the presence of real interviewers may induce the possibility of reasoning and judgment behaviors[1] that are difficult to simulate in a virtual interviewer. Therefore, we expected a significant difference in the behavioral measures of participants as a result of the different self-presentation experience between the virtual interviewer and real interviewer. The hypotheses for these analyses were as follows:

- **H1.1:** The appraisal before the presentation task with the virtual interviewers is not significantly different from the appraisal before the presentation held in front of the human interviewers.

- **H1.2:** The perceived stress elicited by the presentation task with the virtual interviewers is significantly lower than the perceived stress while presenting in front of real interviewers.

- **H1.3:** The perceived difficulty of the presentation task with the virtual interviewers is significantly lower than the perceived difficulty of presenting in front of real interviewers.

- **H1.4:** The behavioral measures of participants with the virtual interviewers are significantly different (significantly higher or lower) than the participants' behaviors with the human interviewers.

The ANOVA results showed that no significant main effects were found (p = 0.91) for Appraisal scores between *Real Interviewer* and *Virtual Interviewer*. This confirms **H1.1**.

The analysis of the perceived stress revealed significant main effects between *Real Interviewer* and *Virtual Interviewer* conditions ($F(1,62) = 8.81$; $p < 0.01$). Lower perceived stress scores were found for the *Virtual Interviewer* condition (Figure 9). This confirms **H1.2**. Regarding the perceived difficulty, no significant main effects were found ($p = 0.28$), thus rejecting **H1.3**.

Regarding the behavioral measures, highly significant mean effects were found for the Contraction Index (CI) between *Real Interviewer* and *Virtual Interviewer* conditions ($F(1,62) = 74.01$; $p < 0.0001$). Higher means were obtained for the Virtual
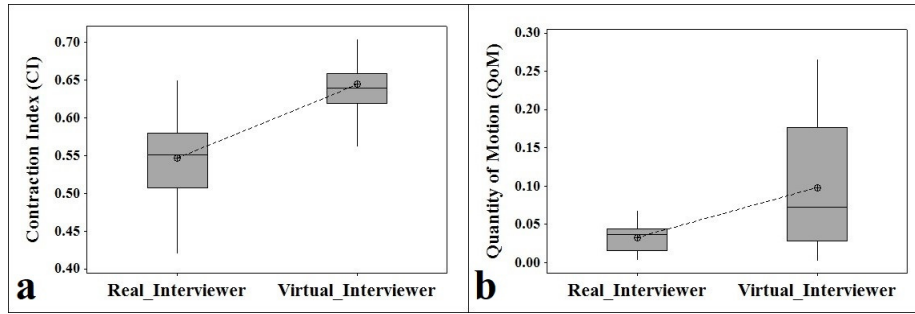
FIGURE 10 Box plots for the Contraction Index (CI) (a) and the Quantity of Motion (QoM) (b). CI was significantly higher for the *Virtual Interviewer* condition. QoM was significantly higher for the *Virtual Interviewer* condition.
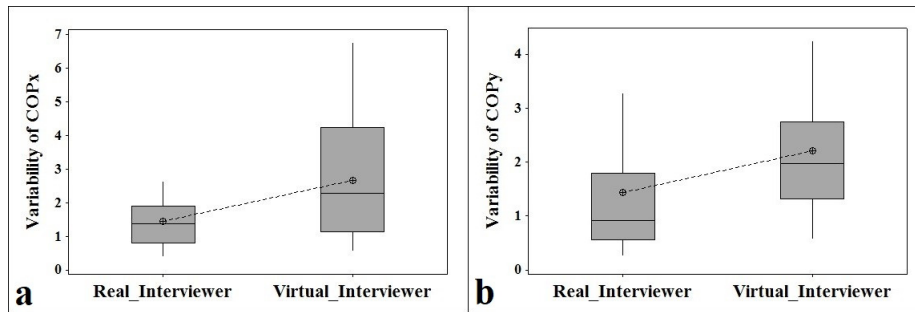

FIGURE 11 Box plots for the variability of ML center of pressure displacement (COP_x) (a) and the variability of the AP center of pressure displacement (COP_y) (b). Participants performed with significantly higher variability (larger sways) in the *Virtual Interviewer* condition.

Interviewer condition (Figure 10a). The Quantity of Motion (QoM) also showed highly significant mean effects ($F(1,62) = 18.90$; $p < 0.0001$), where higher means were found for the Virtual Interviewer condition (Figure 10b). The variability of the center of pressure displacement showed main effects for COP_x ($F(1,62) = 11.24$; $p < 0.01$) and COP_y ($F(1,62) = 6.03$; $p < 0.05$). Both displacements (COP_x and COP_y) showed higher means for the Virtual Interviewer condition (Figure 11). These results confirm **H1.4**.

5.2 | Second analysis: Movement and Gender

Regarding users' perception of the virtual interviewer, a two-factor between-subject ANOVA was conducted for perceived humanness, dominance and warmth. Regarding movement perception, a two-factor between-subjects ANOVA was conducted for each evaluated item: self-movement perception, virtual interviewer movement perception and similarity of the virtual interviewer's movement. Regarding users' emotional responses, a two-factor between-subjects ANOVA was conducted for each emotional dimension: pleasure, arousal and dominance. Regarding the subjective evaluation of the performance, a two-factor between-subjects ANOVA was conducted for perceived difficulty and perceived stress scores. Finally, regarding the behavioral performance, a two-factor between-subjects ANOVA was conducted for each behavioral measure computed from the presentation task: the average of the Contraction Index (CI), the average of the Quantity of Motion (QoM), and the standard deviation of the ML and AP displacements (COP_x and COP_y). Post hoc comparisons were performed using Tukey's method.

As mentioned in section 2.3, several studies[57–59] observed that mimicking or mirroring the postural behavior during a social interaction facilitates the smoothness of interactions between interaction partners. Thus, we expected that the *Movement* factor (*Mirror* / *Random*) would significantly affect the perceived humanness and users' performances (perceived stress and difficulty). Mirroring behavior, by inducing a greater sense of human presence, would induce more humanness and stress than random behavior. Perceived dominance and warmth could be affected by *Gender* since western gender stereotypes associate attributes
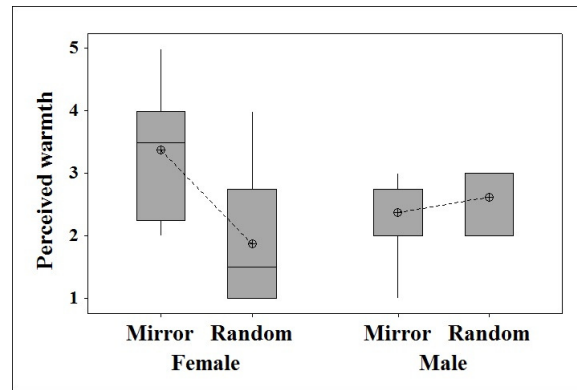
FIGURE 12 Box plot for perceived warmth scores. The mirror movement condition obtained significantly higher mean scores only for the Female virtual interviewer condition.

of affiliativeness and positivity to women, while men are generally perceived as being more dominant and negative[82]. However, since both virtual interviewers wore business clothes with blurred faces, we expected to find no significant differences in either dimension. We expected that the perceived warmth of the virtual interviewer would be affected significantly by the mirror condition: being mirrored should increase the sense of rapport and positivity. We also expected the participants' emotional responses and behaviors to be impacted only by *Movements* based on the idea that mirroring is a spontaneous phenomenon inducing a greater feeling of presence and social stress. This association between perceived humanness, dominance, warmth and perceived stress would allow us to induce higher or lower stress and difficulty in the self-presentation task of the job simulator. Finally, we expected that participants would not notice any difference in the movement of the virtual interviewer. Thus, the hypotheses for this analysis were as follows:

- **H2.1:** Perceived humanness of the virtual interviewer is significantly different only between *Mirror* and *Random* movement conditions.

- **H2.2:** Perceived dominance and warmth are not affected by *Gender*. Perceived warmth is affected significantly by *Movement*.

- **H2.3:** Perceived stress and difficulty during the task are significantly different only between *Mirror* and *Random* conditions.

- **H2.4:** The participants' emotional responses during the task are significantly different only between *Mirror* and *Random* conditions.

- **H2.5:** The behavior of participants during the task is significantly different only between *Mirror* and *Random* conditions.

- **H2.6:** Participants' perceptions of self-movement, virtual interviewer movement and similarity of virtual interviewer movement are not significantly different across all conditions.

Regarding the perceived humanness scores, no significant main effects were found for the *Movement* factor ($p = 0.199$) or the *Gender* factor ($p = 0.63$). The interaction between both factors was found to be non-significant ($p = 0.85$). Thus, hypothesis **H2.1** is rejected.

With respect to the perceived dominance, no significant main effects were found for *Movement* ($p = 0.85$) and Gender ($p = 0.85$). The interaction between factors was also non-significant ($p = 0.560$). Regarding the perceived warmth scores, the analysis showed a non-significant main effect for *Movement* ($p = 0.09$) and *Gender* ($p = 0.729$). However, a significant interaction effect was found between both factors ($F(1,31) = 6.02$; $p < 0.05$). Post hoc analysis revealed that there were significant differences between *Mirror* and *Random* movements only for the female interviewer condition ($p < 0.05$). The *Mirror* movement condition induced higher means of the perceived warmth scores. The *Movement* factor was not significantly different than for the Male condition (Figure 12). These results partly confirm **H2.2**.
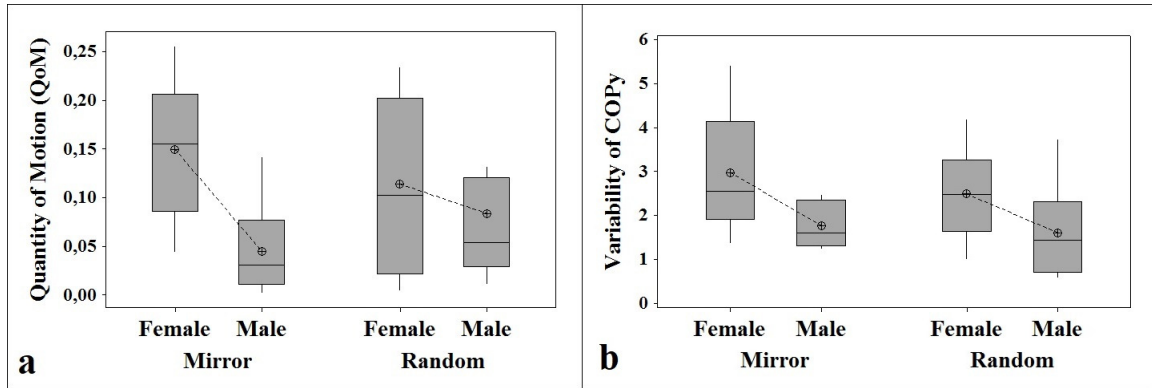
FIGURE 13 Boxplots for Quantity of Motion (QoM) (a) and for variability of the AP center of pressure displacement (COP_y) (b). Participants' QoM values were significantly higher with the *Female* virtual interviewer. A higher difference was obtained with *Mirror* movements. COP_y was significantly higher with the *Female* virtual interviewer.

Perceived stress and difficulty scores did not show significant main effects for *Movement* ($p = 0.88$ for stress, $p = 0.43$ for difficulty) or *Gender* ($p = 0.88$ for perceived stress, $p = 0.79$ for perceived difficulty). No significant interaction effects were found ($p = 0.054$ for perceived stress, $p = 0.07$ for perceived difficulty). This rejects **H2.3**.

Regarding the participant's emotional response (pleasure, arousal and dominance), no significant main effects were found for *Movement* ($p = 0.16$ for pleasure, $p = 0.31$ for arousal, $p = 0.6$ for dominance) or Gender ($p = 0.54$ for pleasure, $p = 0.54$ for arousal, $p = 0.6$ for dominance) factors. The interaction effect was also found to be non-significant ($p = 0.07$ for pleasure, $p = 0.07$ for arousal, $p = 0.38$ for dominance). These results reject **H2.4**.

Regarding the behavior of participants, the Contraction Index (CI) did not show significant main effects for *Movement* ($p = 0.44$) or *Gender* ($p = 0.81$). The interaction among factors was found to be non-significant ($p = 0.46$). The Quantity of Motion (QoM) did not show significant main effects for Movement ($p = 0.96$). Conversely, a significant main effect was found for *Gender* ($F(1,31) = 6.36$; $p < 0.05$). Post hoc tests showed that the Quantity of Motion (QoM) was higher for the Female virtual interviewer condition. Pairwise comparisons showed a higher difference of Quantity of Motion (QoM) for the Mirror condition ($p < 0.05$) (Figure 13a). The interaction between both factors was observed to be non-significant ($p = 0.17$). The variability of the ML center of pressure displacement (COP_x) did not show significant main effects for *Movement* ($p = 0.7$) and *Gender* ($p = 0.15$). In addition, a non-significant interaction was found ($p = 0.28$). Regarding the variability of the AP center of pressure displacement (COP_y), no significant main effects were found for *Motion* ($p = 0.39$). However, the *Gender* factor showed a significant main effect ($F(1,31) = 7.92$; $p < 0.01$). Post hoc tests revealed a higher variability of COP_y in the *Female* virtual interviewer condition (Figure 13b). No significant interaction was found between the two factors ($p = 0.67$). These results partially support **H2.5**.

Regarding the perception of movement by participants, no significant effects were found in the self-movement perception (*Movement* factor: $p = 0.12$, *Gender* factor: $p = 0.34$), the virtual interviewer movement perception (*Movement factor*: $p = 0.72$, *Gender* factor: $p = 1$) or the similarity of virtual interviewer movement (*Movement* factor: $p = 0.29$, *Gender* factor: $p = 0.61$). Interaction between both factors was also observed to not be significant (self-movement perception: p = 0.2, virtual interviewer movement perception: $p = 0.29$, similarity of virtual interviewer movement: $p = 0.1$). These results confirm **H2.6**.

5.3 | Third analysis: Correlations between participants' performances and personality trait

A Pearson's correlation matrix was used to explore the relationship between the performances of participants and their personality traits. Table 3 shows the correlation matrix between the personality traits of participants (**O**penness to experience, **C**onscientiousness, **E**xtraversion, **A**greeableness, and **N**euroticism) and the participants' behavioral measures: Contraction Index (CI), Quantity of Motion (QoM), variability of horizontal center of pressure displacement (COP_x), and variability of vertical center of pressure displacement (COP_y); their emotional response: Pleasure (**P**), Arousal (**A**), and Dominance (**D**); the situation appraisal scores (**Appra**), the perceived stress (**Stress**), and the perceived difficulty (**Diff**).

TABLE 3 Pearson correlation matrix between participants' measures collected during the experiment and their personality traits. Degree of correlation significance: (A) $p < 0.05$, (B) $p < 0.01$, (C) $p < 0.001$

	CI	QoM	COPx	COPy	P	A	D	Appra	Stress	Diff
O	0.091	0.215	0.093	-0.097	0.136	-0.247	**-0.529[B]**	-0.261	-0.290	**-0.401[A]**
C	**0.374[A]**	0.008	**-0.442[A]**	-0.126	0.307	-0.310	**-0.355[A]**	-0.250	**-0.512[B]**	**-0.385[A]**
E	0.306	0.339	0.208	0.125	**0.466[B]**	-0.347	**-0.411[A]**	**-0.416[A]**	-0.183	-0.250
A	0.280	**0.416[A]**	0.094	0.219	0.296	-0.148	-0.157	-0.334	**-0.358[A]**	-0.206
N	-0.084	-0.123	0.027	0.061	**-0.586[C]**	**0.581[C]**	**0.664[C]**	**0.601[C]**	**0.639[C]**	**0.512[B]**
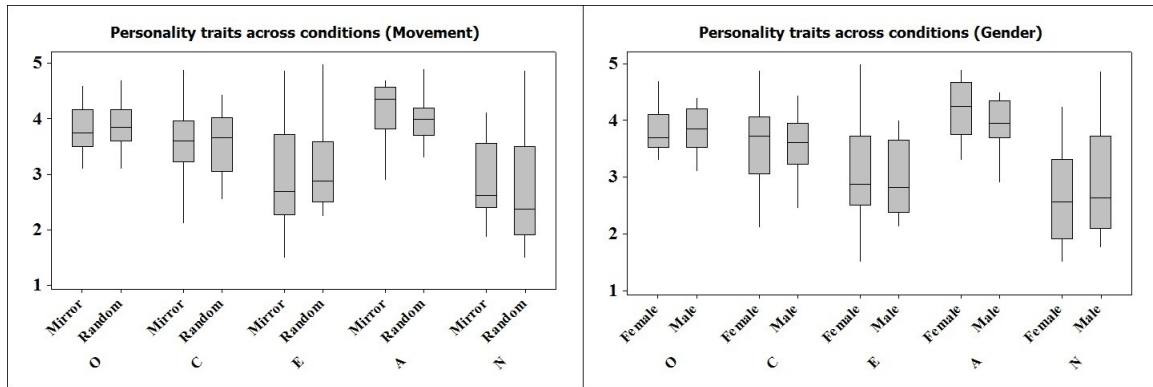
FIGURE 14 Box plot showing the distribution of personality traits (**O**penness to experience, **C**onscientiousness, **E**xtraversion, **A**greeableness, and **N**euroticism) of the participants across conditions (Movement and Gender).

Regarding the behavioral measures, the correlation analyses (Table 3) revealed a weak and positive correlation between the Contraction Index and the Conscientiousness personality trait ($r = 0.374$, $p < 0.05$). The Quantity of Motion was moderately and positively correlated to the Agreeableness personality trait ($r = 0.416$, $p < 0.05$). The variability of COP_x was moderately and negatively correlated to Conscientiousness ($r = -0.442$, $p < 0.05$).

Regarding the emotional response, Pleasure was moderately and positively correlated to Extraversion ($r = 0.466$, $p < 0.01$) and moderately and negatively correlated to Neuroticism ($r = -0.586$, $p < 0.001$). Arousal was moderately and positively correlated to Neuroticism ($r = 0.581$, $p < 0.001$). Dominance was moderately and negatively correlated to Openness to experience ($r = -0.529$, $p < 0.01$) and Extraversion ($r = -0.411$, $p < 0.05$). It was also weakly and negatively correlated to Conscientiousness ($r = -0.355$, $p < 0.05$) and strongly and positively correlated to Neuroticism ($r = 0.664$, $p < 0.001$).

Appraisal scores were strongly and positively correlated to Neuroticism ($r = 0.601$, $p < 0.001$) and moderately and negatively correlated to Extraversion ($r = -0.416$, $p < 0.05$). Perceived Stress was strongly and positively correlated to Neuroticism ($r = 0.639$, $p < 0.001$), moderately and negatively correlated to Conscientiousness ($r = -0.512$, $p < 0.01$) and weakly and negatively correlated to Agreeableness ($r = -0.358$, $p < 0.05$). Perceived Difficulty was moderately and positively correlated to Neuroticism ($r = 0.512$, $p < 0.01$), moderately and negatively correlated to Openness to experience ($r = -0.401$, $p < 0.05$) and weakly and negatively correlated to Conscientiousness ($r = -0.385$, $p < 0.05$).

Finally, the distribution of personality traits of the participants across conditions is showed in Figure 14. As shown in this figure, the personality traits with more variability were Neuroticism and Extraversion.

6 | DISCUSSION

Our results suggest that the human-virtual agent interactive situation was found to be as threatening as the human-human situation *before* the interaction and as difficult but with lower perceived stress *after* the interaction. This lower perceived stress with virtual interviewers may be due to the fact that the feedback session (*dialogue phase*) with the interviewers was not included in

this interaction. Thus, the participants were not evaluated about their performance by the virtual interviewer. However, although the perceived stress was lower in the interaction with the virtual interviewer compared to the interaction with the human interviewer, the absolute ratings on the stress scale for both situations were close, revealing an average level of stress (about the mean of the scale). These results suggest that it is relevant for users to train with a virtual simulation of a job interview that reproduces the conscious appraisal process of a real-life training situation (simulation of a job interview with real persons). The difference in stress felt between the real and virtual conditions is congruent with most of the differences in behaviors: being less stressed induces less controlled behaviors (more quantity of motion; center of pressure displacements that correspond to a more articulated mode of intersegmental functioning, freeing joint degrees of freedom). In fact, as negative or threat stimuli can trigger flee behaviors, fight behaviors or freezing behaviors[83], a job interview is considered to be a situation that can induce freezing behaviors[19]. The increase in the contraction index for the virtual interviewer simulation is more difficult to interpret. One possible explanation is that, although candidates felt less stressed and had larger postural movements, inducing a greater quantity of movement, they displayed less hand movements (which increased the contraction index) because of the virtual nature of the interviewer: communicative gestures are less necessary when interacting with such an artificial entity, the hands of which are not visible. Such a result and interpretation require further analyses, for example, by comparing hand movements of the participants from the real and virtual conditions via video analysis. Overall, the fact that our virtual simulation is less stressful than the human-human simulation was expected: it seems challenging to elicit as much stress as the presence of a real person. Future studies should attempt to evaluate how the presence of richer nonverbal information (facial expression, head and hand movements) displayed by the virtual interviewer can reduce this gap.

We compared female vs. male virtual interviewer and mirror vs. random postural movements conditions. We expected no effect of *Gender* on the simulation since clothes and faces were not salient indicators of gender. This result was partially confirmed: no main effect of *Gender* was found. Again, this research question remains open for situations where facial expressions are visible: western stereotypes are possibly reduced by the limited amount of gendered cues. An additional factor potentially limiting any gender effect is the fact that motion-captured movements were from a male only. Regarding mirroring, the condition was subtle and not detected by participants. This was confirmed by self-reports of perceived movement similarity. Such a result is important since mirroring has been observed to not be well perceived by users if it is too salient and easily detectable[57]. Second, we hypothesized that mirroring as a spontaneous phenomenon would induce a greater feeling of presence associated with a more stressful interview. This hypothesis was not validated: no differences in rated humanness or in perceived stress or difficulty and postural behaviors were observed. Instead, an unexpected interaction appeared to be significant: the female virtual interviewer under the mirroring condition was rated as being warmer, inducing a greater quantity of movements and more postural displacements in the vertical plane. red This effect may not be related with the gender of participants since the gender of the thirty-two participants was also balanced in our experiment (sixteen males and sixteen females). This effect seems to be related to an increase of rapport thanks to a congruence effect: the female virtual interviewer, which conveys more positivity than the male interviewer, induces a significantly more pleasant interaction when mirroring, thus leading to a perception of affiliation[84].

The correlations between participants' states and behavior measures and personality traits were computed. However, as the sample size is relatively small, the results should be taken cautiously. Globally, the results suggest that the inter-individual differences in personality factors are in relation to: postural behaviors, emotional reactions and perceptions of participants during the simulated presentation task. Regarding behavioral measures, Agreeableness was observed to be positively correlated with quantity of motion. This is in accordance with the impact of an increase of rapport in the interaction: more agreeable participants had more affiliative behaviors, leading to an increase in Quantity of Motion[85]. More conscious participants were more contracted (contraction index) and displayed fewer lateral postural movements. Such a result is in line with the fact that *Conscientiousness* is related to being more careful in decisions but also in behaviors[86]. Regarding self-reported variables, *Neuroticism* is directly associated with self-esteem, itself being tied to the manner of appraising others' judgments. Being more neurotic leads to low self-esteem, which leads to increased fear of rejection and to social anxiety[63]. As such, this trait was related to all other reported variables in an expected way: being more neurotic induced less pleasure, more arousal and more dominant emotions (e.g., anger, anxiety, frustration, etc.), appraising the situation as more threatening, stressful and difficult. The fact that no behavioral measure is related to this trait was not expected: interaction effects could be investigated in the future. *Openness* and *Extraversion* (i.e., positive dispositions) are also known to have an influence on stress response through an effect on the appraisal process[63]. While both were negatively related to perceived threat, only the correlation with *Extraversion* was significant. In addition, *Extraversion* (which is related to more emotional positivity) was related to more pleasure, and *Openness* was related to a perceived less

difficult task. With such a multidimensional association, increasing the sample size to allow multivariate analysis (e.g., canonical correlations) would allow the drawing of firmer conclusions.

7 | CONCLUSIONS AND FUTURE WORK

In this paper, we detailed how we designed a job interview simulator that replicates a job interview with real humans (the Trier Social Stress Test[9]. Our contribution to the relevant literature lies in our focus on postural movements of seated interviewers and standing candidates. We focused on the presentation phase of the job interview: the virtual interviewer is listening, the face blurred, and leaving only postural movements visible for the candidate (no facial expressions). The job interview simulator was evaluated via two analyses: a comparison between the virtual job simulator and a collected corpus with real human interviewers and a comparison between different characteristics of virtual interviewers (*Gender* and *Movement*). Finally, an exploratory analysis was conducted to investigate the correlation between the participants' performances during the presentation task and their personality traits.

Overall, the user study revealed the feasibility of a virtual simulation of a job interview that reproduces at least the conscious appraisal process of a real-life training situation via postural movements only. Characteristics of virtual interviewers (*Gender* and *Mirroring*) and personality traits of participants impacted the experience, confirming the possibility of providing a more personalized job interview simulator.

To extend the current work, several directions can be pursued. First, we would like to integrate additional non-verbal cues in the virtual interviewers' behaviors (non-blurred face, facial expressions, head nods, gestures) to enhance the simulated presentation task. Second, the postural interaction can be improved by combining multiple users' postural cues (quantity of motion, contraction index, 3D posture, gender-specific biological movements, etc.) to continuously adapt and synchronize the postural behavior of the virtual interviewer to the user's postural behavior. Third, we would like to analyze the impact of the direction of postural moves of the virtual interviewer under the random condition (e.g., a virtual interviewer constantly leaning back in his/her chair could be perceived by the users as being uninterested). Fourth, we would like to adapt the current interaction into a Virtual Reality setup to improve the immersion of our system. In this way, we can compare if there are significant differences regarding the perceived stress and difficulty, emotional responses and behavior of participants between the VR simulation and the current setup with the screen. Fifth, we will expand the set of postural moves of the virtual interviewer in order to increase the realism in the movements of the virtual interviewer. Finally, it would be interesting to evaluate the usefulness of the proposed simulator by conducting a longitudinal study with multiple sessions to see if the users' improvements induced by the virtual training transfer to a job interview with human interviewers.

References

1. Harmon-Jones E, and Winkielman P. Social Neuroscience: Integrating Biological and Psychological Explanations of Social Behavior. Guilford Press; 2007.

2. Anderson K, André E, Baur T, Bernardini S, Chollet M, Chryssafidou E, et al. The TARDIS Framework: Intelligent Virtual Agents for Social Coaching in Job Interviews. In: Proceedings of the 10th International Conference on Advances in Computer Entertainment - Volume 8253. ACE 2013. Berlin, Heidelberg: Springer-Verlag; 2013. p. 476–491. Available from: https://doi.org/10.1007/978-3-319-03161-3_35.

3. Hoque ME, Courgeon M, Martin JC, Mutlu B, and Picard RW. MACH: My Automated Conversation Coach. In: Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing. UbiComp '13. New York, NY, USA: ACM; 2013. p. 697–706.

4. Kleinsmith A, and Bianchi-Berthouze N. Affective Body Expression Perception and Recognition: A Survey. IEEE Transactions on Affective Computing. 2013 Jan;**4**(1):15–33.

5. Delaherche E, Chetouani M, Mahdhaoui A, Saint-Georges C, Viaux S, and Cohen D. Interpersonal Synchrony: A Survey of Evaluation Methods across Disciplines. IEEE Transactions on Affective Computing. 2012 July;**3**(3):349–365.

6. Lazarus RS. Progress on a cognitive-motivational-relational theory of emotion. The American psychologist. 1991;**46**:819–34.

7. Lazarus RS. Stress and emotion: A new synthesis. 2007 01;p. 33–51.

8. Kudielka B, Hellhammer D, and Kirschbaum C. In: Ten years of research with the Trier Social Stress Test—revisited; 2007. p. 56–83.

9. Kirschbaum C, Pirke KM, and Hellhammer DH. The 'Trier Social Stress Test'–a tool for investigating psychobiological stress responses in a laboratory setting. Neuropsychobiology. 1993;**28**:76–81.

10. Mejias H, Applebaum RL, Applebaum SJ, and Trotter RT. Oral communication apprehension and Hispanics: An exploration of oral communication apprehension among Mexican American students in Texas. In:) EKHDJYE, editor. Language anxiety: From theory and research to classroom implications. Englewood Cliffs, NJ:Prentice-Hall; 1991. p. 87 – 97.

11. Behnke RR, and Sawyer CR. Anticipatory anxiety patterns for male and female public speakers. Communication Education. 2000;**49**(2):187–195.

12. Gaibani A, and Elmenfi F. The Role of Gender in Influencing Public Speaking Anxiety. Int J Gend Women's Stud. 2014;**2**:105–116.

13. Matsuda S, and Gobel P. Anxiety and predictors of performance in the foreign language classroom. System. 2004;**32**(1):21 – 36.

14. Kane EW, and Macaulay LJ. Interviewer Gender and Gender Attitudes. The Public Opinion Quarterly. 1993;**57**(1):1–28.

15. Tourangeau R, Couper MP, and Steiger DM. Humanizing self-administered surveys: experiments on social presence in web and IVR surveys. Computers in Human Behavior. 2003;**19**:1–24.

16. Streeter LA, Macdonald NH, Apple W, Krauss RM, and Galotti KM. Acoustic and perceptual indicators of emotional stress. The Journal of the Acoustical Society of America. 1983;**73**(4):1354–1360.

17. Roelofs K, Hagenaars MA, and Stins J. Facing Freeze: Social Threat Induces Bodily Freeze in Humans. Psychological Science. 2010;**21**(11):1575–1581.

18. Wallbott HG, and Scherer KR. Stress specificities: differential effects of coping style, gender, and type of stressor on autonomic arousal, facial expression, and subjective feeling. Journal of personality and social psychology. 1991;**61 1**:147–56.

19. Giraud T, Jáuregui DAG, Hua J, Isableu B, Filaire E, Scanff CL, et al. Assessing Postural Control for Affect Recognition Using Video and Force Plates. In: 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction; 2013. p. 109–115.

20. Baggett HL, Saab PG, and Carver CS. Appraisal, Coping, Task Performance, and Cardiovascular Responses During the Evaluated Speaking Task. Personality and Social Psychology Bulletin. 1996;**22**(5):483–494.

21. Ramanarayanan V, Leong CW, Chen L, Feng G, and Suendermann-Oeft D. Evaluating Speech, Face, Emotion and Body Movement Time-series Features for Automated Multimodal Presentation Scoring. In: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. ICMI '15. New York, NY, USA: ACM; 2015. p. 23–30.

22. Aigrain J, Dapogny A, Bailly K, Dubuisson S, Detyniecki M, and Chetouani M. On Leveraging Crowdsourced Data for Automatic Perceived Stress Detection. In: Proceedings of the 18th ACM International Conference on Multimodal Interaction. ICMI 2016. New York, NY, USA: ACM; 2016. p. 113–120. Available from: http://doi.acm.org/10.1145/2993148.2993200.

23. Aigrain J, Spodenkiewicz M, Dubuisson S, Detyniecki M, Cohen D, and Chetouani M. Multimodal stress detection from multiple assessments. IEEE Transactions on Affective Computing. 2018;p. 1–1.

24. Nguyen LS, and Gatica-Perez D. I Would Hire You in a Minute: Thin Slices of Nonverbal Behavior in Job Interviews. In: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. ICMI '15. New York, NY, USA: ACM; 2015. p. 51–58.

25. Tanveer MI, Zhao R, Chen K, Tiet Z, and Hoque ME. AutoManner: An Automated Interface for Making Public Speakers Aware of Their Mannerisms. In: IUI; 2016. .

26. Naim I, Tanveer MI, Gildea D, and Hoque ME. Automated Analysis and Prediction of Job Interview Performance. IEEE Transactions on Affective Computing. 2016 April;**9**(2):191–204.

27. Kang N, Brinkman W, van Riemsdijk MB, and Neerincx MA. An Expressive Virtual Audience with Flexible Behavioral Styles. IEEE Transactions on Affective Computing. 2013;**4**(4):326–340.

28. Vanni FV, Conversano C, Debbio AD, Landi P, Carlini MP, Fanciullacci C, et al. A survey on virtual environment applications to fear of public speaking. European review for medical and pharmacological sciences. 2013;**17 12**:1561–8.

29. Pertaub D, Slater M, and Barker C. An Experiment on Public Speaking Anxiety in Response to Three Different Types of Virtual Audience. Presence. 2002 Feb;**11**(1):68–78.

30. Batrinca LM, Stratou G, Shapiro A, Morency LP, and Scherer S. Cicero - Towards a Multimodal Virtual Audience Platform for Public Speaking Training. In: IVA; 2013. p. 116–128.

31. Jones H, Sabouret N, Damian I, Baur T, André E, Porayska-Pomsta K, et al. Interpreting social cues to generate credible affective reactions of virtual job interviewers. In: International Workshop on Intelligent Digital Games for Empowerment and Inclusion. Haifa, Israel; 2014. .

32. Gebhard P, Baur T, Damian I, Mehlmann G, Wagner J, and André E. Exploring Interaction Strategies for Virtual Characters to Induce Stress in Simulated Job Interviews. In: Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems. AAMAS '14. Richland, SC; 2014. p. 661–668.

33. DeVault D, Artstein R, Benn G, Dey T, Fast E, Gainer A, et al. SimSensei Kiosk: A Virtual Human Interviewer for Healthcare Decision Support. In: Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems. AAMAS '14. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems; 2014. p. 1061–1068.

34. Chollet M, Stefanov K, Prendinger H, and Scherer S. Public Speaking Training with a Multimodal Interactive Virtual Audience Framework. In: Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. ICMI '15. New York, NY, USA: ACM; 2015. p. 367–368.

35. Wörtwein T, Chollet M, Schauerte B, Morency LP, Stiefelhagen R, and Scherer S. Multimodal Public Speaking Performance Assessment. In: ICMI; 2015. .

36. Philip P, Dupuy L, Auriacombe M, Serre F, de Sevin E, Sauteraud A, et al. Trust and acceptance of a virtual psychiatric interview between embodied conversational agents and outpatients. npj Digital Medicine. 2020 12;**3**.

37. Morency LP, de Kok I, and Gratch J. A probabilistic multimodal approach for predicting listener backchannels. Autonomous Agents and Multi-Agent Systems. 2010 Jan;**20**(1):70–84.

38. Poppe R, Truong KP, and Heylen D. Perceptual evaluation of backchannel strategies for artificial listeners. Autonomous Agents and Multi-Agent Systems. 2013 Sep;**27**(2):235–253.

39. Huang L, Morency LP, and Gratch J. Virtual Rapport 2.0. In: Vilhjálmsson HH, Kopp S, Marsella S, and Thórisson KR, editors. Intelligent Virtual Agents. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011. p. 68–79.

40. de Kok I, Ozkan D, Heylen D, and Morency LP. Learning and Evaluating Response Prediction Models Using Parallel Listener Consensus. In: International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction. ICMI-MLMI '10. New York, NY, USA: ACM; 2010. p. 3:1–3:8.

41. Wang Z, Lee J, and Marsella S. Towards More Comprehensive Listening Behavior: Beyond the Bobble Head. In: IVA; 2011. .

42. Buschmeier H, and Kopp S. When to Elicit Feedback in Dialogue: Towards a Model Based on the Information Needs of Speakers. In: Bickmore T, Marsella S, and Sidner C, editors. Intelligent Virtual Agents. Cham: Springer International Publishing; 2014. p. 71–80.

43. Schroder M, Bevacqua E, Cowie R, Eyben F, Gunes H, Heylen D, et al. Building Autonomous Sensitive Artificial Listeners. IEEE Transactions on Affective Computing. 2012 April;**3**(2):165–183.

44. Bull P. Posture and Gesture. vol. 16. Pergamon Books; 1987.

45. Coulson M. Attributing Emotion to Static Body Postures: Recognition Accuracy, Confusions, and Viewpoint Dependence. Journal of Nonverbal Behavior. 2004 Jun;**28**(2):117–139.

46. Tan N, Clavel C, Courgeon M, and Martin JC. Postural Expressions of Action Tendencies. In: Proceedings of the 2Nd International Workshop on Social Signal Processing. SSPW '10. New York, NY, USA: ACM; 2010. p. 53–58.

47. Winter DA. Biomechanics and Motor Control of Human Movement. Hoboken, N.J. : John Wiley Sons; 2009.

48. Duarte M, and Freitas SMSF. Revision of Posturography Based on Force Plate for Balance Evaluation. Brazilian Journal of Physical Therapy. 2010 06;**14**:183 – 192.

49. Horslen BC, and Carpenter MG. Arousal, valence and their relative effects on postural control. Experimental Brain Research. 2011 Sep;**215**(1):27.

50. Lang PJ, Davis M, and Öhman A. Fear and anxiety: animal models and human cognitive psychophysiology. Journal of Affective Disorders. 2000;**61**(3):137 – 159. Arousal in Anxiety.

51. Frijda NH. The Emotions. University Press.; 1986.

52. Hagenaars MA, Oitzl M, and Roelofs K. Updating freeze: Aligning animal and human research. Neuroscience Biobehavioral Reviews. 2014;**47**:165 – 176. Available from: http://www.sciencedirect.com/science/article/pii/S0149763414001869.

53. Stins JF, Roelofs K, Villan J, Kooijman K, Hagenaars MA, and Beek PJ. Walk to me when I smile, step back when I'm angry: emotional faces modulate whole-body approach–avoidance behaviors. Experimental Brain Research. 2011 Aug;**212**(4):603–611.

54. Condon WS, and Osgton WD. Speech and body motion synchrony of the speaker-hearer. In: Horton D, and Jenkins J, editors. The perception of language. New York: Academic Press; 1971. p. 150–184.

55. La France M, and Ickes W. Posture mirroring and interactional involvement: Sex and sex typing effects. Journal of Nonverbal Behavior. 1981 Mar;**5**(3):139–154.

56. Guéguen N, Jacob C, and Martin A. Mimicry in Social Interaction: Its Effect on Human Judgment and Behavior. European Journal of Social Sciences – Volume Number. 2009 04;**8**.

57. Lakin JL, and Chartrand TL. Using Nonconscious Behavioral Mimicry to Create Affiliation and Rapport. 2003 08;**14**:334–9.

58. Yabar Y, Johnston L, Miles L, and Peace V. Implicit Behavioral Mimicry: Investigating the Impact of Group Membership. Journal of Nonverbal Behavior. 2006;**30**:97–113.

59. Chartrand T, and Bargh J. The chameleon effect: the perception-behavior link and social interaction. Journal of personality and social psychology. 1999;**76 6**:893–910.

60. Shockley KD, Santana M, and Fowler CA. Mutual interpersonal postural constraints are involved in cooperative conversation. Journal of experimental psychology Human perception and performance. 2003;**29 2**:326–332.

61. Shockley KD, Richardson DC, and Dale R. Conversation and Coordinative Structures. Topics in cognitive science. 2009;**1 2**:305–19.

62. Giraud T, Soury M, Hua J, Delaborde A, Tahon M, Jauregui DAG, et al. Multimodal Expressions of Stress during a Public Speaking Task: Collection, Annotation and Global Analyses. In: 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction; 2013. p. 417–422.

63. Schneider TR, Rench TA, Lyons JB, and Riffle RR. The Influence of Neuroticism, Extraversion and Openness on Stress Responses. Stress Health. 2012;**28**:102–110.

64. Bobick AF, and Davis JW. The recognition of human movement using temporal templates. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2001 March;**23**(3):257–267.

65. Kret M, and de Gelder B. When a smile becomes a fist: the perception of facial and bodily expressions of emotion in violent offenders. Experimental Brain Research. 2013;**228**:399 – 410.

66. Proverbio A, Ornaghi L, and Gabaro V. How face blurring affects body language processing of static gestures in women and men. Social Cognitive and Affective Neuroscience. 2018;**13**:590 – 603.

67. Gelder B, de Borst A, and Watson R. The perception of emotion in body expressions. Wiley Interdisciplinary Reviews: Cognitive Science. 2015 03;**6**:149–158.

68. Gu Y, Mai X, and Luo Y. Do Bodily Expressions Compete with Facial Expressions? Time Course of Integration of Emotional Signals from the Face and the Body. PLoS ONE. 2013;**8**.

69. Technologies U. Unity3D. [Online; accessed 02-April-2019]; 2020. https://unity.com/.

70. Wagner J, Lingenfelser F, Baur T, Damian I, Kistler F, and André E. The social signal interpretation (SSI) framework: multimodal signal processing and recognition in real-time. In: ACM Multimedia; 2013. .

71. Plaisant O, Courtois R, and Réveillère C;.

72. Bonardi A, Truck I, and Akdag H. Building Fuzzy Rules in an Emotion Detector. In: IPMU 2006 - 11th International Conference on Information Processing and Management of Uncertainty. Paris, France; 2006. p. 540–546.

73. Castellano G, Villalba SD, and Camurri A. Recognising Human Emotions from Body Movement and Gesture Dynamics. In: Paiva ACR, Prada R, and Picard RW, editors. Affective Computing and Intelligent Interaction. Berlin, Heidelberg: Springer Berlin Heidelberg; 2007. p. 71–82.

74. Microsoft. Kinect SDK. [Online; accessed 04-April-2020]; 2020. https://developer.microsoft.com/en-us/windows/kinect.

75. Giraud T, Focone F, Demulier V, Martin JC, and Isableu B. Perception of Emotion and Personality Through Full-Body Movement Qualities: A Sport Coach Case Study. ACM Trans Appl Percept. 2015 Oct;**13**(1):2:1–2:27.

76. Srivastava SM, and John OP. The Big-five Trait Taxonomy: History, Measurement, and Theoretical Perspectives; 1999. .

77. Chartrand TL, and Bargh JA;.

78. Ho CC, and MacDorman KF. Revisiting the Uncanny Valley Theory: Developing and Validating an Alternative to the Godspeed Indices. Comput Hum Behav. 2010 Nov;**26**(6):1508–1518.

79. Mori M. The uncanny valley. Energy. 1970;**7**(4):33–35.

80. Buzzotta VR, Lefton RE, and Sherberg M. Effective Selling Through Psychology: Dimensional Sales and Sales Management. New York, NY.: Wiley Interscience; 1972.

81. Bradley MM, and Lang PJ. Measuring emotion: The self-assessment manikin and the semantic differential. Journal of Behavior Therapy and Experimental Psychiatry. 1994;**25**(1):49 – 59.

82. Brody LR, and Hall J. Handbook of Emotions. In: Lewis M, and Haviland-Jones J, editors. Gender, emotion, and expression. 2nd ed. Guilford Press; 2000. p. 338–349.

83. Hagenaars MA, Stins JF, and Roelofs K. Aversive life events enhance human freezing responses. Journal of experimental psychology General. 2012;**141**(1):98–105.

84. BERNIERI FJ. dyad rapport and accuracy of its judgment across situations : a lens model analysis. Journal of Personality and Social Psychology. 1996;**71**(1):110–129. Available from: https://ci.nii.ac.jp/naid/30016756822/en/.

85. Kang SH, Gratch J, Wang N, and Watt JH. Agreeable People Like Agreeable Virtual Humans. In: Prendinger H, Lester J, and Ishizuka M, editors. Intelligent Virtual Agents. Berlin, Heidelberg: Springer Berlin Heidelberg; 2008. p. 253–261.

86. DePaulo BM. Nonverbal behavior and self-presentation. Psychol Bull. 1992;**111**(2):203–243.

AUTHOR BIOGRAPHY

David Antonio Gómez Jáuregui is an Assistant Professor of Computer Science at ESTIA Institute of Technology (Bidart, France). His research interests are focused on creating new ways of natural user interaction with computers by detecting emotions, gestural and non-verbal behaviours from users. This is a multi-disciplinary research involving Human-Computer Interaction, Affective Computing, Computer Vision and Artificial Intelligence. Before joining ESTIA, he was as a postdoctoral researcher at IBM France Center for Advanced Studies (2015-2016). He was also a postdoctoral researcher at LIMSI-CNRS (2012-2015) and at INRIA Rennes (2011-2012). He received his PhD in Computer Science at Télécom SudParis, France, on May 2011. He is supervising 1 PhD thesis and 1 post-doc.

Tom Giraud is a postdoc in Human Computer Interaction at the Human-Centered Multimedia Lab, University of Augsburg. His researches focus on the design of mixed interfaces engaging the sensitive body – embodied interaction – around embedded installations – situated interaction.

Brice Isableu is a Full Professor in Cognitive Psychology and Ergonomics at Aix Marseille University, ALLSH. He heads the department of Cognitive and experimental psychology at Aix Marseille University. He conducts his research at PSYCLE laboratory (Cognition, Emotion, Langage) at Aix-en-Provence where he heads a team of researchers in Cognitive Psychology and Ergonomics on the topics of "Contextualized activities and ergonomics". His research domain focused on the study of real, virtual and augmented multimodal interactions (social and affective and motor) in Humans and in Machines/Systems, with a special emphasis on associated motor signatures. His research goals aims both at 1) better understanding and identifying factors (cognitive, conative, sensorimotor, social, affective processes) of inter and intra-individual differences observed at various levels and scales of human-human and Humans-machine/systems interactions (avatars, robots, virtual reality) and 2) exploiting results from 1) to develop personalized computational systems, assistance systems and methods taking into account users preferences (in healthy or pathological subjects) and their associated adaptive capabilities to facilitate multimodal interactions with virtual reality or agents and humanoid robots during demanding dynamic and uncertain activities. User preferences are studied in young and healthy subjects, during aging, in pathological subjects, and during the development of sports expertise. These questions are addressed using psychophysical methods, statistical Bayesian formalism, embodied and situated cognition approach, 3D movement modeling. He supervised 11 defended PhD theses and 2 post-doc.

Jean-Claude Martin is 1st class full-time Professor of Computer Science at Université Paris Saclay, France. He conducts his research at LIMSI-CNRS where he heads a team of researchers in Psychology and in Computer Sciences. His research domain is Social and Affective Multimodal Interaction in Humans and in Machines. His research goals are two-folds: 1) inspire from theories and models from Psychology to design human-machine interactions, and 2) design human-machine interactions which enable to better understand humans. He works in several application areas related to social skills training and eHealth: autism ; job and

medical interviews ; virtual coaches ; leadership and teamwork ; stress management ; sports. He is Editor-in-Chief of the Springer Journal on Multimodal User Interfaces (2018 IF = 2.020). He supervised 15 defended PhD theses plus 4 which are on-going.

A | ANNEX: MOVEMENT PERCEPTION QUESTIONNAIRE

I moved during the interaction.

Never		1.☐	2.☐	3.☐	4.☐	5.☐	Often

The virtual interviewer moved during interaction.

Never		1.☐	2.☐	3.☐	4.☐	5.☐	Often

If he (she) moved, its movements were the same as mine.

Disagree		1.☐	2.☐	3.☐	4.☐	5.☐	Totally agree