Perception of Avatars Non-verbal Behaviours in Virtual Reality

ETIENNE Elodie[1], LECLERCQ Anne-Lise[2], REMACLE Angélique[3], DESSART Laurence[4], and SCHYNS Michaël[5]

[1]QuantOM, HEC Liège, N1 UER Opérations, University of Liège, Liège, Belgique
[2]Research Unit for a life-Course perspective on Health & Education, Speech-Language Pathology Department, B38B Logopédie Clinique, University of Liège, Liège, Belgique
[3]Research Unit for a life-Course perspective on Health & Education, Speech-Language Pathology Department, B38B Logopédie des Troubles de La Voix, University of Liège, Liège, Belgique
[4]Strategic Marketing Innovation, HEC Liège, N1 Marketing et intelligence stratégique, University of Liège, Liège, Belgique
[5]QuantOM, HEC Liège, N1 UER Opérations: Informatique de Gestion, University of Liège, Liège, Belgique

Highlights

- The findings guide the design of avatars for training frontline staff in virtual reality to improve their interactions with customers.
- Using photo-realistic avatars increases confidence in assessing valence and arousal levels.
- Using simple cardboard is effective for assessing the valence and arousal with a sufficient level of confidence.
- High-end virtual reality headsets enhance the sense of presence, creating a more realistic training experience.
- A more nuanced interpretation reveals that the smiles of avatars are predominantly perceived as neutral in this study.

Abstract

Virtual reality has shown great potential in many fields, especially in business and psychology. By immersing someone in a new computer-generated reality, it is possible to create realistic, safe, and controllable simulations for research and training, as well as new three-dimensional-enriched consumer experiences and services. Most of these environments, especially in the metaverse, rely on virtual representations of people called "avatars". The design and non-verbal behaviours of these avatars must be carefully crafted to provide a realistic and truly immersive experience. This paper aims to understand how avatar non-verbal behaviours (i.e., body posture, facial expression, and head movement) are perceived by users immersed in a virtual reality context, a very common situation encountered in many simulations and especially during training. Therefore, the first objective of this study is to validate, through an experiment with 125 participants, how the audience's levels of emotional valence and arousal are perceived in virtual reality. Based on these results, a library of audience non-verbal behaviours corresponding to different arousal and valence levels is now available for future applications. The experiment also examines the benefits of using low-end versus high-end virtual reality headsets, and photo-realistic versus cartoon avatars. The results have implications for the design of realistic, challenging, and interactive virtual audiences.

Keywords: Virtual reality, non-verbal behaviour, avatars, presence, frontline employees, speech performance, communication skills, training

1 Introduction

Virtual reality offers tremendous opportunities in business practice due to its ability to provide realistic, safe, and controllable experiences and services. By immersing individuals in a life-like environment and replicating real-life situations (Remacle et al., 2021), virtual reality is particularly valuable when human behaviour is a critical element of the business process under study. Not surprisingly, virtual reality research is especially prevalent in the fields of marketing and psychology. Loureiro et al. (2019) studied no less than 150 papers in 115 journals on the use of virtual reality in marketing. In a special issue on virtual reality in marketing, Boyd and Koles (2019) provide essential insights into various aspects of virtual reality and its implications for this field. Virtual reality, but also augmented reality, are now widely used in many business areas, such as tourism (Beck et al., 2019; Jingen Liang and Elliot, 2021), retail (Bonetti et al., 2018), and health (Howard et al., 2021; Howard, 2017 and van Bennekom et al.,2017. Virtual reality is also widely used in education through training (Vlachopoulos and Makri,2017; Hainey et al. 2016; Remacle et al.;2023).

A concept related to virtual reality is the *metaverse*, which Hennig-Thurau et al. (2022, p. 1) define as a "computer-mediated environment consisting of virtual worlds, within which users can act and communicate in real-time using virtual people". In these virtual worlds, if virtual people (also called *avatars*) are present, their design and non-verbal behaviours must be realistic to enhance the learning experience. Avatars can be defined as "digital entities with anthropomorphic appearance, controlled by a human or software, that have an ability to interact" (Miao et al., 2022, p. 67). Accordingly, the term avatar is used in this study to refer to interactive representations of people, whether realistic or cartoons. The metaverse and avatars are becoming increasingly popular concepts, especially in marketing. For instance, in their recent study, Dwivedi et al. (2023) argue that the metaverse has the potential to revolutionise the way businesses interact with consumers. The authors discuss various applications of the metaverse in marketing, including brand experiences, customer engagement, and immersive advertising, highlighting the challenges and ethical considerations associated with using the metaverse in marketing, such as privacy concerns and potential biases in avatar design.

The present study focuses on public speaking situations where an avatar listens to a user speaking. Public speaking is a widespread activity in everyday life. Whether in front of a small or large audience, individuals often need to mobilise their public speaking skills. Psychology research has investigated different activities, such as presenting research findings, teaching and lecturing in classrooms or professional development workshops (Remacle et al., 2021; Remacle et al., 2023). There are also more complex situations that involve public speaking, such as providing psycho-education to patients in individual or group therapy settings, advocating for mental health policies and initiatives at public events or rallies, leading support groups for individuals with mental health challenges, and so forth (Freeman et al., 2017).

Focusing on public speaking tasks in various fields and a virtual environment with an interactive virtual audience (i.e., providing non-verbal feedback) has been the focus of several studies (Chollet and Scherer, 2017; Chollet et al., 2016a; Chollet et al., 2016b; Batrinca et al., 2013; Chollet et al., 2021; Kang et al., 2016; Glémarec et al., 2019; Lugrin et al., 2016; Seufert et al., 2022). Virtual reality allows public speakers to practice in a situation very similar to what they will face in real life. People can practice where, when, and as often as they need to build their confidence and be ready to give a speech (Scheveneels et al., 2019; Craske et al., 2014). In addition, the training can be progressive and controlled in terms of the number of participants, their non-verbal behaviours, and so on, resulting in much more efficient and faster learning (Eckert and Mower, 2020).

Focusing on public speaking tasks in various fields and a virtual environment with an interactive virtual audience (i.e., providing non-verbal feedback) has been the focus of several studies (Chollet and Scherer, 2017; Chollet et al., 2016a; Chollet et al., 2016b; Batrinca et al., 2013; Chollet et al., 2021; Kang et al., 2016; Glémarec et al., 2019; Lugrin et al., 2016; Seufert et al., 2022). Virtual reality allows public speakers to practice in a situation very similar to what they will face in real life. People can practice where, when, and as often as they need to build their confidence and be ready to give a speech (Scheveneels et al., 2019; Craske et al., 2014). In addition, the training can be progressive and controlled in terms of the number of participants, their non-verbal behaviours, and so on, resulting in much more efficient and faster learning (Eckert and Mower, 2020). There are also a number of tasks in marketing that involve public speaking: pitching a product or service to potential customers or investors, giving a progress report to a team or supervisor, presenting a proposal or idea to a group or board of directors, and interacting with clients. The practical application of this study is improving the public speaking skills of frontline employees who play a critical role in fostering positive consumer perceptions through effective communications in face-to-face interactions. Frontline employees play a crucial role in delivering high-quality services that increase customer satisfaction and loyalty (Maria Stock et al., 2017). In such customer-facing situations, frontline employees would benefit from public speaking training. However, providing effective training for these employees can be a challenging task, particularly in the context of rapidly evolving service technologies and complex service encounters (Larivière et al., 2017; Robinson et al., 2020, Dwivedi et al., 2023). Virtual reality has also emerged as a promising technology for employee training, providing an immersive and interactive learning experience that can improve their skills, knowledge, and confidence in dealing with customers. Indeed, virtual reality technology has become increasingly popular in frontline employee training across a range of industries (De Keyser et al., 2019) due to its ability to simulate real-life situations, allowing trainees to develop and practice their skills in a safe and controlled environment.

Whilst the benefits of virtual reality for training purposes have been demonstrated, blind spots remain in the precise application and design of this technology to maximise training effectiveness and foster positive perceptions among trainees. Specifically, the trainee-avatar interaction is relevant to understanding the impact that the listening avatar in the simulated environment may have on the trainee practising public speaking. Indeed, the immersed person must perceive these listening avatars as realistic in order to provide a truly immersive and effective experience. However, the impact of the avatar's perceived responses on the trainee and the resulting sense of presence in the virtual reality environment remain unclear. Furthermore, the development of accurate and authentic non-verbal avatar behaviour is still in its infancy. Therefore, the first objective of this study is to investigate how virtual reality users perceive the levels of emotional valence and arousal of avatars. The second and third objectives concern a more in-depth analysis of the sense of presence using different types of headsets and different avatar qualities. In sum, this study aims to investigate how avatars are perceived and the benefits of using different headsets or graphics.

2 Theoretical Development

One of the key benefits of virtual reality in training frontline employees is its ability to simulate real-life service encounters and practice situational skills. According to Lloyd and Luk (2011), interaction behaviours that lead to feeling comfortable in the service encounter are crucial to creating a positive customer experience. Virtual reality can help trainees develop these behaviours by creating realistic simulations of different service encounters, allowing them to practice their skills in a safe and controlled environment. This can help build confidence and reduce anxiety in real-life service encounters. In addition, virtual reality can provide a platform for role-playing exercises that are effective in developing the skills and behaviours required for successful service encounters. Solomon et al. (1985) suggest that role theory can be used to understand the dynamics of service encounters, with service employees and customers playing specific roles in the interaction. Virtual reality can provide a platform for employees to practice different roles and scenarios, allowing them to develop a better understanding of the service encounter dynamics and how to respond effectively. Furthermore, virtual reality can help improve the quality of service encounters by enhancing employees' non-verbal communication skills. According to Sundaram and Webster (2000), non-verbal communication plays a critical role in service encounters, with cues such as facial expressions, gestures, and tone of voice conveying important information to customers. Virtual reality can also be used to simulate various non-verbal cues, allowing employees to practice their communication skills and become more effective in conveying information to customers. As the question of how frontline employees respond to consumer emotions in service encounters is still a black hole (Rafaeli et al., 2017), virtual reality can help address this question.

In addition, virtual reality can provide a cost-effective and scalable solution for training frontline staff (Hilken et al., 2022). Farrell et al. (2001) suggest that conceptualizing service encounters is important for understanding employee behaviour and how it affects customer perceptions of service quality. However, traditional training methods, such as classroom teaching and on-the-job training, can be time consuming and costly. Virtual reality can provide a cost-effective solution for training large numbers of employees, allowing organizations to scale their training programs without incurring high costs. Deng et al. (2019) highlight the importance of technology infusion in frontline service delivery. They argue that technological advances have changed the nature of service encounters and created new opportunities for service providers to improve the customer experience. However, they also note that effectively integrating technology in service encounters requires adequate training and support for frontline staff. Virtual reality can be an effective tool for such training by simulating realistic service encounters in a controlled environment. However, research on the appropriate design of a virtual reality environment for frontline employee training is still lacking (Khandelwal and Upadhyay, 2021).

In the context of public speaking in general, virtual reality can improve communication self-efficacy (Remacle et al., 2023) and speaking performance (Chollet and Scherer, 2017), especially when the audience (the listening avatar) is interactive (Wörtwein et al., 2015; Goberman et al., 2011). The avatar's reactions can have a significant impact on the speaker's emotions and performance (Pertaub et al., 2002; Poeschl, 2017; Kleinsmith et al., 2015). As shown in Chollet et al. (2015) and Chollet et al. (2021), interactivity in virtual reality is also an important part of the training process. Therefore, it is essential to know if the user perceives the interactions between the avatar and her/himself in the virtual environment as representative of reality and how each interaction is interpreted. Two dimensions related to emotion and affect are essential: the level of arousal and emotional valence. According to Chollet and Scherer (2017) *arousal* can be understood as the avatar's level of alertness, while *valence* corresponds to how positively or negatively the avatar feels toward the speaker or the presentation. Specifically, body postures, facial expressions, and head movements are found to express some degrees of arousal and valence (Argyle and Ingham, 1972; Knapp et al., 2013).

In the context of service encounters, marketers face the challenge of creating positive experiences that meet or exceed customer expectations. To achieve this goal, marketers can use the concepts of valence and arousal, which are critical in shaping customer perceptions and behaviours during service encounters. A positive emotional response to a service encounter can lead to more favourable perceptions of the service provider and the service itself, resulting in higher customer satisfaction and loyalty (Babin and Attaway, 2000; Brady et al., 2001). However, negative emotional responses and excessive arousal can have the opposite effect, resulting in lower customer satisfaction and loyalty. Negative emotional responses can lead to negative word-of-mouth , which can have a significant impact on a service provider's reputation (Smith and Wheeler, 2002). In addition, excessive arousal can lead to anxiety and discomfort, which can negatively impact the overall experience and result in lower customer satisfaction (Liao and Chuang, 2004). Valence and arousal are important concepts in marketing, particularly in the context of service encounters. They play a critical role in shaping customer perceptions and behaviours, thereby influencing their satisfaction and loyalty. By designing service encounters that promote positive emotional experiences and appropriate levels of arousal, service providers can increase customer satisfaction, loyalty, and ultimately business success. It is therefore important to gain a better understanding of how frontline employees perceive the emotional valence and arousal of avatars in virtual reality.

This leads to two hypotheses linking emotional valence and arousal to body postures, facial expressions, and head movements.

H1: Avatar smiles and nods are associated with positive valence by users, frowns and head shakes with negative valence, and raised eyebrows are mostly neutral.

H2: Higher levels of avatar facial expressions, head movements, and forward postures are positively related to higher levels of arousal perceived by participants.

Following Slater's (2003, p. 1) terminology, *immersion* is what the technology delivers from an objective point of view, or the objective level of sensory fidelity a virtual reality system provides. *Presence* is the human response to immersion, i.e., the participant's subjective sense of being in the virtual place. It is the "observer's sense of psychologically leaving their real location and feeling as if transported to a virtual environment" (Weech et al. 2019, p. 2). It has long been known that presence can affect user performance (Nash et al., 2000). For a virtual reality environment and its associated goals to be effective (in this case, training frontline employees to adequately respond to the emotions of avatarised consumers), the presence and the immersion components of virtual reality must be as high as possible.

To study variations in presence due to the immersive quality of the virtual reality environment, the device used to display the virtual reality is of paramount importance (Flavián et al., 2019; Leveau and Camus , 2023). Indeed, low-end virtual reality headsets, i.e., smartphones in Samsung Gear VR Cardboards (see Figure 1), and high-end virtual reality headsets, i.e., Meta Rift S headsets (see Figure 2) are expected to lead to significant differences in the sense of presence. Due to hardware limitations, the low-end headset is not expected to perform better than the high-end headset (higher display resolution, wider field of view, more accurate tracking, amongst others), and the latter is often the preferred choice in terms of sense of presence (e.g., Selzer et al., 2019). The reasons for this are cost, ease of operation, and portability. As Amin et al. (2016) explain in comparing immersion in the Cardboard and Oculus Rift headsets, a high-end headset is expensive, requires professional technical operation, and is less mobile than Cardboard with its affordability (only a smartphone is needed in addition to the Cardboard) and ease of use. A high-end headset can thus make virtual reality inaccessible to everyone or enable mass adoption (a hundred people at a time). Amin et al.'s, 2016 results suggest that Cardboard is capable of providing an acceptable level of immersion. Similar results may be derived when assessing valence and arousal between low-end and high-end headsets. Should this be the case, it may be possible to address the challenges of affordability and widespread adoption at a lower cost. This leads to the third hypothesis regarding the device used.

Figure 1: Low-end headset: Samsung Gear Virtual Reality headset

Figure 2: High-end headset: Oculus Rift S headset

H3:

a. Both headsets, including the low-end, are capable enough to accurately evaluate emotional valence and arousal with confidence.

b. High-end headsets provide a higher sense of presence than low-end headsets.

The third objective concerns the quality of the avatars used. In the event that virtual reality and the metaverse come to fruition, avatars will assume the role of our digital representations or embodiments (Kim et al., 2023). Most often the virtual audience is represented by "simplified" cartoon avatars (in terms of the method used to create them), but interactive three-dimensional avatars have the highest level of the digital object continuum, as Koles and Nagy (2021) explain. Several studies examine the importance to users of the creation of their avatars (Sibilla and Mancini, 2018; Schrader, 2019; Kang and Kim, 2020; Seymour et al., (2018). For example, Seymour et al., 2018 examine whether participants have different affinity, trustworthiness[1], and preferences for avatars with two levels of realism (one almost human-realistic and one a cartoon). Another study (Seymour et al., 2021) shows that participants who interacted with the human-realistic avatar in a two-dimensional video display had a positive experience, rated the avatar as more trustworthy, had more affinity, and preferred it. The level of confidence in their answers was therefore higher when they dealt with human-realistic avatars. When the same questions were

[1]Trust can be applied to information systems as explained in Wang and Benbasat (2005))

asked of participants immersed in virtual reality, the affinities and preferences were even stronger. Similar results were found for robots and their anthropomorphism. For example, Letheren et al. (2021) found in their work that consumers prefer a higher level of humanness and a moderate to high level of sociability. Other studies emphasise the importance of anthropomorphism in promoting trust and positive responses (Kim et al., 2019; Van Pinxteren et al., 2019; Murphy et al., 2019; Fraune et al., 2020; Belanche et al., 2021; Yao et al., 2023). Regarding the comparison of avatar non-verbal behaviours, Krumhuber et al. (2009) show a high degree of correspondence between the effects produced by synthetic and human smiles. In addition, Cafaro et al. (2016) highlight that first impressions may determine important relationship decisions. The final question therefore examines whether fully rigged three-dimensional photo-realistic models can significantly improve participants' perceptions of the avatar's arousal and valence or their confidence level. The desire for this comparison lies in the ease of avatar creation. In the context of a virtual audience, it would be possible to create a virtual environment with avatars that look like the people supposedly present in the real context (Schrader, 2019). The last hypothesis is therefore related to the method used to create the avatars.

H4: Compared to cartoon avatars, photo-realistic avatars improve the level of confidence of users.

3 Methodology

3.1 The Virtual Reality Environment

The avatars and the environment were created by three-dimensional animation specialists, technical artists, and three-dimensional game artists working in a research lab at our university. The technical artists used the Unity three-dimensional engine. The three-dimensional artists used the *Blender* and *Maya* software. The photo-realistic avatars were created and animated using the *Reallusion* suite.

The virtual reality environment (see Figure 3) depicts an office with the avatar behind a desk, animated (see Section 3.2 for the possible animations) as he listens to the immersed participant for at least 15 seconds. The environment was designed to represent a common interaction between a frontline employee and a customer, such as in the context of a banking, insurance, or real estate agency appointment. The environment was optimised to support the highest quality virtual reality experience possible using an autonomous low-end Samsung S7 smartphone and a high-end Oculus headset connected to a computer (HP Omen) with a virtual reality-compatible graphic power unit (NVIDIA GeForce GTX1070 graphics card) and processor (Intel(R) Core(TM) i7-8550U CPU @ 1.80GHz 1.99 GHz). Due to the relative simplicity of the scene, it was not necessary to drastically reduce the complexity of the three-dimensional avatars.

Figure 3: Screenshot of the virtual reality environment used for the experiment

Eight avatars were created for this experiment, represented in Figure 4 in their neutral state in the Unity software. Thus, in the virtual reality environment, these images represent the quality of the avatars, not their non-verbal behaviours. To determine whether the method used to create the avatars influenced participants' perceptions of the avatar's arousal and valence or their interpretation of the audience's confidence levels, half were cartoon avatars and half were photo-realistic avatars based on real people. Photo-realistic avatars are created from photos of real people (women and men) who gave informed consent. The software used is *Character Creator* with the plugin *HeadShot*, which is part of the *Reallusion* suite. The avatars were then enhanced by three-dimensional artists and animated by three-dimensional animators. For the sake of diversity and to avoid gender bias, half of the avatars are female and half male. In terms of origin, four avatars are European, two are Indian with a dark complexion, and two are African.

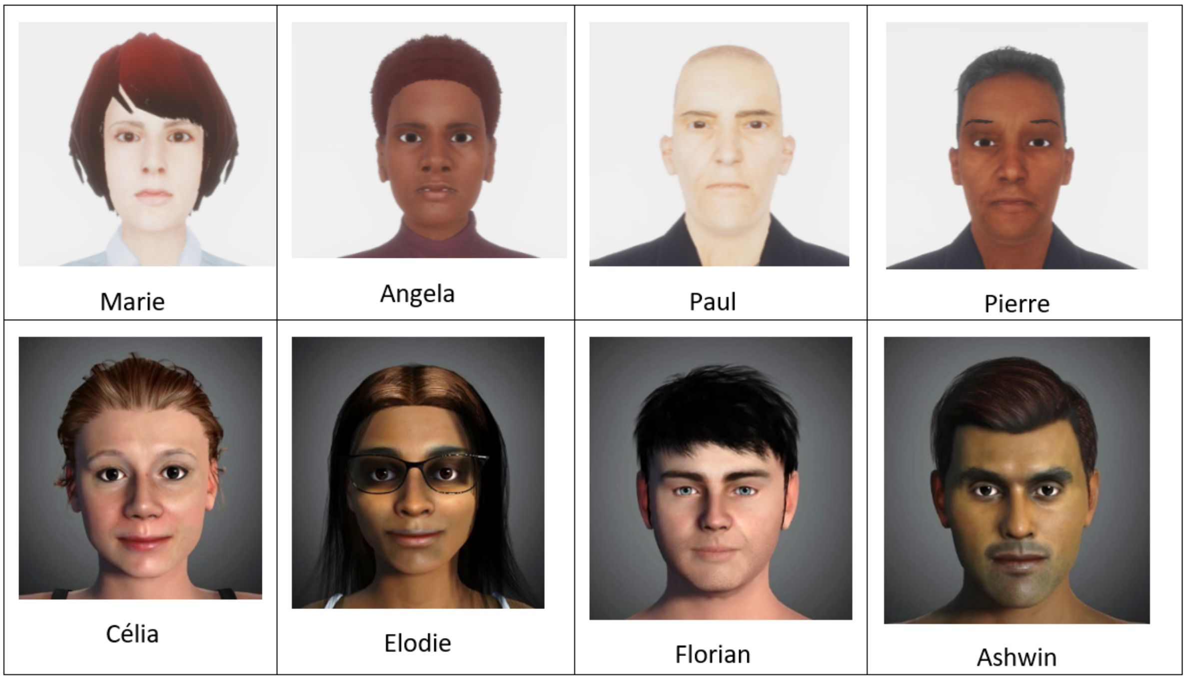

Figure 4: Avatars created using two different methods. The first row represents cartoon avatars and the second row photo-realistic avatars.

3.2 Non-Verbal Behaviours

The avatars and their animations were created by three-dimensional and technical artists. Several animations were created for different types of avatars (male or female avatars of European, African, or Indian origin). To define the sets of animations, the results of Chollet and Scherer (2017) were used as a basis. The parameters considered in their study include: posture (forward, backward, neutral), amount of averted gaze (0, 25, 50, 75, or 100%), direction of averted gaze (sideways, down, or up), type of facial expression (smile, frown, or raised eyebrows), facial expression frequency, head movements (nod or shake), and head movement frequency. Considering all these parameters for a true immersion virtual reality experiment is excessive and unnecessary, since the will is not to study the valence and level of arousal for each possible combination. In fact, only the most common non-verbal behaviours of an audience are needed. Therefore, the results of Chollet and Scherer (2017) were used to define appropriate combinations that are expected to represent positive, neutral, and negative valence with a low or high level of arousal. For example, a head nod is expected to have positive valence in contrast to a head shake, and a forward posture is expected to have a higher level of arousal. Table 1 displays the parameters used. It leads to 144 possible combinations (9 postures x 4 facial expressions x 4 head movements).

In order to reduce the number of manageable combinations in the experiment, a set of 40 non-verbal behaviours (as described in the next section) was randomly selected. In this selection process, a priori *clear* animations (e.g., smile and nod), neutral animations, and more ambiguous animations (e.g., smile and shake) were intentionally kept to broadly cover the set of possibilities.

3.3 The Virtual Reality Conditions

An experiment was conducted to determine whether a set of predefined animated avatars were interpreted in virtual reality to express different levels of valence and arousal. Each participant had to rate the level of valence and arousal for 20 animated 15-second sequences. Across participants, 40 sequences (i.e., combination of postures, facial expressions, and head movements) were tested. Table 2 shows these animations.

The avatars and animations were divided into four sets[2]. Each set contains one male and one female avatar, one photo-realistic version and one cartoon version. Ten animations were randomly attached to each of these two avatars.

[2]Throughout this paper, a distinction is made between the animations (sequences), which are divided into *sets*, and participants divided into *groups*.

Table 1: Parameters of non-verbal behaviour

Postures (P)	**Facial expressions (F)**	**Head (H)**
1: Backward posture – Arms crossed 2: Backward posture –Arms stand (elbows on the table with hands crossed) 3: Backward posture – Arms behind head 4: Upright posture – Hand on hand (hands on the table, one on top of the other) 5: Upright posture – Hands together (hands crossed on the table) 6: Upright posture – Hands separated 7: Forward posture – Hands together 8: Forward posture – Arms stand (elbows on the table with hands crossed) 9: Froward posture – Arms crossed	1: None 2: Smiling 3: Frowning 4: Eyebrows raised	1: None 2: Nod 3: Shake 4: Questioning

Table 2: Sequences

Seq. 1	Seq. 2	Seq. 3	Seq. 4	Seq. 5	Seq. 6	Seq. 7	Seq. 8	Seq. 9	Seq. 10
P1F3H4	P3F4H2	P7F2H2	P7F2H4	P7F3H3	P7F4H3	P7F4H4	P9F1H2	P9F2H3	P9F3H3
Seq. 11	Seq. 12	Seq. 13	Seq. 14	Seq. 15	Seq. 16	Seq. 17	Seq. 18	Seq. 19	Seq. 20
P2F2H3	P2F4H2	P3F3H3	P3F4H3	P4F4H1	P4F4H3	P5F3H1	P7F3H1	P8F4H1	P9F1H3
Seq. 21	Seq. 22	Seq. 23	Seq. 24	Seq. 25	Seq. 26	Seq. 27	Seq. 28	Seq. 29	Seq. 30
P2F1H2	P2F4H4	P3F1H2	P3F3H2	P3F4H4	P4F1H2	P5F1H4	P6F3H4	P7F1H1	P7F1H2
Seq. 31	Seq. 32	Seq. 33	Seq. 34	Seq. 35	Seq. 36	Seq. 37	Seq. 38	Seq. 39	Seq. 40
P1F3H1	P2F4H3	P3F2H3	P3F4H1	P4F1H3	P4F2H2	P6F3H3	P7F1H4	P7F4H3	P8F3H4

Each participant in the experiment viewed 20 sequences from one of the four sets. The sequences were randomly presented. For the first three sets, all the experiments were conducted using an autonomous low-end virtual reality headset based on a Samsung S7 smartphone and the Samsung Gear virtual reality adapter, an improved version of the Google Cardboard. The experiment was repeated for the Set 3 sequences using a high-end Oculus headset connected to a computer with an Nvidia Geforce GTX1070 graphics card. The experiment for Set 4 was also conducted using a high-end Oculus headset connected to a computer with an Nvidia Geforce GTX1070 graphics card.

3.4 The Experiment Settings

The present study was approved by the Ethics Committee of the Faculty of Psychology, Speech Therapy, and Educational Sciences of a European University (file number: 1920-81). The participants were recruited voluntarily through a targeted advertisement sent via the official university e-mail. A website was created and participants had to register and provide informed consent after receiving a full description of the study. Once registered, they were given an anonymous identifier that allowed them to access the online questionnaire once the experiment began. Based on a power test, 24 participants were required to rate each animation in order to draw valid conclusions. 125 participants took part in the experiment. The participants were immersed in a virtual reality environment to rate the valence and arousal of a single avatar. They were divided into 5 groups by order of registration[3]. To test the third hypothesis (a and b), the same animation settings were used for the third and fourth groups, but with different headsets.

After being given the definitions of valence and arousal (written at the beginning of their online questionnaire) as defined in Section 2, participants were asked to wear the provided headset (Cardboard and Samsung S7, or Rift S headset). The virtual reality application contained the list of animations to be viewed. The 20 animated sequences were presented in random order. Each participant had to watch each animation for at least 15 seconds (at the end of the 15 seconds, the sequence repeated endlessly until the participant moved on), and each animation was played in its entirety before moving on to the next one. In addition, the sequence was restarted as often as necessary. For each of the 20 animations, using their computer or smartphone, they responded to an online

[3]25 participants for the first group with the Samsung Gear, 25 in the second group with the Samsung Gear, 25 in the third group with the Samsung Gear, 25 in the third group with the Oculus, 25 for the fourth group with the Oculus. The first three groups used the Samsung Gear, i.e., the low-end headset, and the two other groups the Oculus, i.e., the high-end headset.

questionnaire set up for this experiment. Participants had to remove the headset after each sequence to complete the questionnaire. They were asked to rate their level of arousal (7-point Likert-type scale from very low to very high) and valence (7-point Likert-type scale from very negative to very positive), as well as their level of confidence for each of their answers (7-point Likert-type scale from very unconfident to very confident). They could also write down open-ended comments about the video they had watched.

After watching all the sequences (post-immersion), participants answered some questions (in the online questionnaire) about the sense of presence they experienced in the virtual reality environment. The French-Canadian version of The Gatineau Presence Questionnaire (Laforest et al., 2016) was used (Bouchard and Robillard, 2019). It is a 4-item questionnaire scored on a percentage scale. As Laforest et al., (2016, p. 4) explain, Gatineau Presence Questionnaire consists of four items: "1) the feeling of being there, 2) the perception of the experience as real, 3) the awareness of the virtual environment as artificial, and 4) the feeling of being in the physical office instead of a virtual environment". The last two items were scored in reverse, and the average percentage was computed to obtain the global score for the Gatineau Presence Questionnaire and thus the sense of presence in general. In the following, these items will be summarised as 1) presence in virtual reality, 2) level of realism, 3) level of artificiality, and 4) spatial awareness. At the end of the experiment, participants completed the online sociodemographic questionnaire, reporting any recent events that might have altered their perception, such as an upsetting or joyful event.

A total of 125 people (64 men and 58 women) participated in the study, with an age range between 17 and 62 (with an average of 27.25), and most are Belgian (with French as mother tongue).

4 Results

This section analyses emotional valence and level of arousal for the parameters of the non-verbal behaviours separately (posture, facial expression, and head movement). Then, the sequences presented in Table 2 are analysed for emotional valence and level of arousal. The two subsequent sections compare the use of the low-end and high-end headsets, and photo-realistic versus cartoon avatars. Finally, the results of the Gatineau Presence Questionnaire are presented.

4.1 Analysis of Posture, Facial Expression, and Head Movement

This section first presents the levels of arousal and valence attributed by participants for each parameter (posture, facial expression, and head movement). The global goal is to identify in the list of possibilities the behaviours that are clearly associated with specific levels of valence (positive, neutral, negative) and arousal (high, neutral, low). An association is *clear* when all participants, or at least the vast majority, interpret a behaviour in the same way. If the interpretations differ significantly among participants, special care must be taken when using them. A Chi-square test was performed to see if the participants' responses were equally distributed between negative interpretations (responses 1 to 3 in the Likert-type scale) and positive interpretations (responses 5 to 7 in the Likert-type scale)[4]. The aim is to determine whether the difference between the proportions of positive and negative interpretations is statistically significant. For the sake of brevity, Table 3 summarises the results of the statistical tests. The full analysis can be found in the Appendix (see Tables A1 and A2).

In order to determine whether some postures are more significant than others in assessing valence or arousal, ANOVA tests were conducted. Only significant results are presented in this section. Among the backward postures, the crossed arms posture is considered the most negative (p-value $= 0.041$). Similarly, among the head movements associated with negative valence, head shaking expresses the most negative valence. Furthermore, although arousal is almost always perceived as positive, the degree of positivity differs between postures, facial expressions, and head movements (p values < 0.001).

4.2 Library of Animated Avatars

A particularly interesting question is how the combination of the parameters (postures, facial expressions, and head movements) is perceived. Some combinations may enhance the arousal or valence perception, or conversely, blur the results. The perception of the different sequences presented in Table 2 is next analysed. Figures 5 and 6 show the results for emotional valence and arousal. Each row in the figure corresponds to a particular sequence, where the distribution of responses is shown according to the legend below the figure (responses from 1 to 7 represent the 7 points of the Likert-type scale). The three percentages represent the proportion of negative responses (1 to 3), neutral responses (4), and positive responses (5 to 7). The sequences are ordered by the percentage of positive responses (decreasing) and then by the percentage of negative responses (increasing). The first row is thus the most positive sequence for the parameter assessed, and the last is the most negative. For example, for emotional valence, Sequence 36 (P4F2H2: an avatar with an upright posture, hands on top of each

[4]Throughout this paper, and for the sake of clarity, all responses from 1 to 3 in the Likert-type scale are referred to as "negative" and all responses from 5 to 7 in the Likert-type scale are referred to as "positive". However, the words "negative" and "positive" need to be used with care. For valence, it can be understood as negative and positive levels, for arousal, as low and high levels.

Table 3: Interpretations of the Chi-square tests

		Interpretation valence	Interpretation arousal
Postures			
P1	Backward posture – Arms crossed	-	+
P2	Backward posture – Arms down	-	+
P3	Backward posture – Arms behind the head	-	+
P4	Upright posture – Hand on hand	/	+
P5	Upright posture – Hands together	-	/
P6	Upright posture – Hands separated in front	-	+
P7	Forward posture – Hands together	/	+
P8	Forward posture – Arms down	-	/
P9	Forward posture – Arms crossed	-	+
Facial expressions			
F1	Neutral	/	+
F2	Smiling	/	+
F3	Frowning	-	+
F4	Eyebrows raised	-	+
Head movements			
H1	Neutral	-	+
H2	Nod	+	+
H3	Shake	-	+
H4	Questioning	-	+

other, a neutral facial expression, and nodding his head) is perceived as having the most positive emotional valence, while Sequence 14 (P3F4H3: an avatar with a backward posture, hands behind his head, frowning eyes, and shaking his head) is perceived as having the most negative emotional valence (see Figure 5). For level of arousal, Sequence 11 (P2F2H3: an avatar with a backward posture, elbows on the table, a smiling face, and shaking his head) is perceived as having the highest level of arousal, whereas Sequence 30 (P7F1H2: an avatar with a forward posture, hands together, a neutral face, and nodding the head) is perceived as having the lowest level of arousal (see Figure 6).

To determine which sequence is associated with a particular level of valence and arousal, further tests were conducted (using the same method as explained in Section 4.1), and three groups emerged (see Figures 5 and 6, and Table 4):

- Sequences for which the difference between the proportion of negative and positive responses is statistically significant and for which there are more positive than negative responses. These are the sequences above the green line in the figures.
- Sequences for which the distribution of the Likert-type scale responses are equally distributed between negative responses and positive responses. These are the sequences between the green and the red line in the figures.
- Sequences for which the difference between the proportion of negative and positive responses is statistically significant and for which there are more negative than positive responses. These are the sequences below the red line in the figures.

Comparing the results from the previous section (see Table 3) and this section (see Figures 5 and 6) shows that sometimes non-verbal behaviours are separately perceived in same way, but their combination leads to a different perception. For example, regarding emotional valence, in Sequence 4 (i.e., P7F2H4: an avatar with a forward posture, with elbows on the table, head tilted), all the non-verbal behaviours were separately perceived as neutral or negative, but their combination led to positive valence. In this case, a smile seems to be perceived as positive and outperforms the evaluation of other non-verbal behaviours. Another example of the level of arousal is Sequence 29 (i.e., P7F1H1: an avatar with a forward posture, without any facial expression or head movement). In this sequence, all the non-verbal behaviours separately were perceived as representing a high level of arousal. However, their combination led to a perception of low arousal.

A closer look at the categories in Table 4 shows that some non-verbal behaviours dominate over others. For example, every time an avatar shakes his head (sequences associated with the third head movement, i.e., P*F*H3), it is perceived as having negative emotional valence and a high level of arousal, regardless of the associated posture or facial expression. Similarly, when the avatar nods his head (sequences associated with the second head movement, i.e., P*F*H2), it is mainly perceived as having positive emotional valence. Sometimes, depending on the other parameters (posture and facial expression), it is perceived as neutral, but never as negative, as expected. This confirms the first hypothesis.

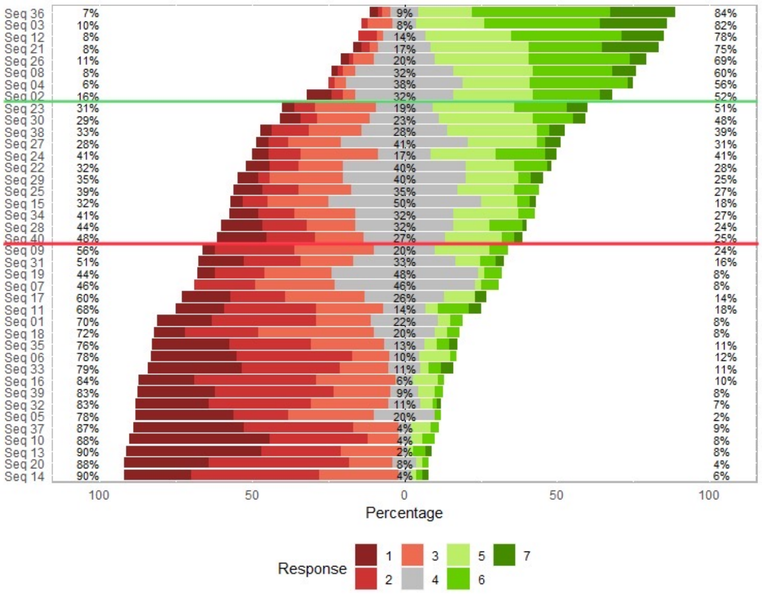

Figure 5: Valence per sequence evaluated on a seven-point Likert scale. All sequences above the green line are perceived as positive, and all sequences under the red line are perceived as negative.

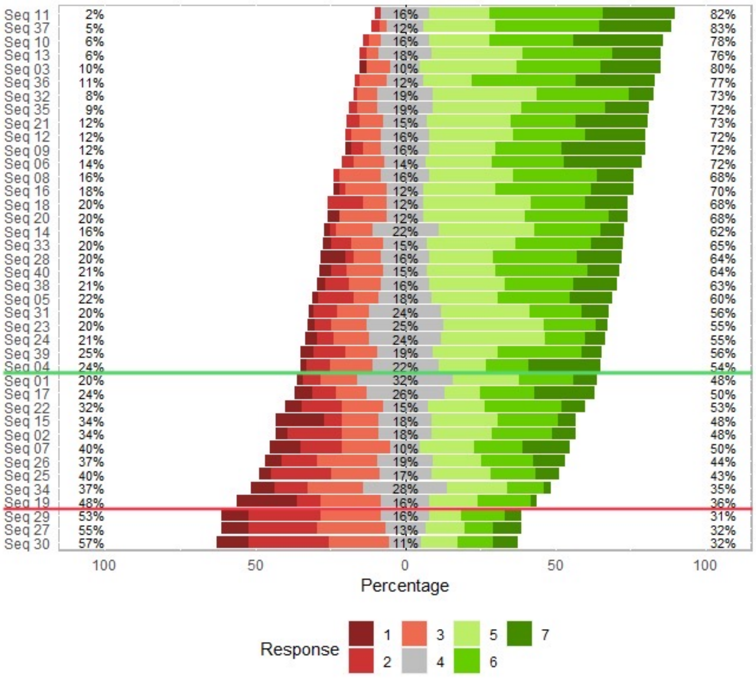

Figure 6: Arousal per sequence evaluated on a seven-point Likert scale. All sequences above the green line are perceived as positive, and all sequences under the red line are perceived as negative.

Seq. 1	Seq. 2	Seq. 3	Seq. 4	Seq. 5	Seq. 6	Seq. 7	Seq. 8	Seq. 9	Seq. 10
P1F3H4	P3F4H2	P7F2H2	P7F2H4	P7F3H3	P7F4H3	P7F4H4	P9F1H2	P9F2H3	P9F3H3
Seq. 11	Seq. 12	Seq. 13	Seq. 14	Seq. 15	Seq. 16	Seq. 17	Seq. 18	Seq. 19	Seq. 20
P2F2H3	P2F4H2	P3F3H3	P3F4H3	P4F4H1	P4F4H3	P5F3H1	P7F3H1	P8F4H1	P9F1H3
Seq. 21	Seq. 22	Seq. 23	Seq. 24	Seq. 25	Seq. 26	Seq. 27	Seq. 28	Seq. 29	Seq. 30
P2F1H2	P2F4H4	P3F1H2	P3F3H2	P3F4H4	P4F1H2	P5F1H4	P6F3H4	P7F1H1	P7F1H2
Seq. 31	Seq. 32	Seq. 33	Seq. 34	Seq. 35	Seq. 36	Seq. 37	Seq. 38	Seq. 39	Seq. 40
P1F3H1	P2F4H3	P3F2H3	P3F4H1	P4F1H3	P4F2H2	P6F3H3	P7F1H4	P7F4H3	P8F3H4

Postures	**Facial expressions**	**Head**
1: Backward posture – Arms crossed	1: None	1: None
2: Backward posture – Arms down	2: Smiling	2: Nod
3: Backward posture – Arms behind head	3: Frowning	3: Shake
4: Upright posture – Hand on hand	4: Eyebrows raised	4: Questioning
5: Upright posture – Hands together		
6: Upright posture – Hands separated		
7: Forward posture – Hands together		
8: Forward posture – Arms stand		
9: Froward posture – Arms crossed		

Recall of Tables 1 and 2 for the sake of readability

Table 4: Sequences per level of valence and arousal
(Please refer to Tables 1 and 2 for the meaning of the parameters and sequences)

	Negative valence	Neutral valence	Positive valence
Low level of arousal	∅	Seq. 27: P5F1H4 Seq. 29: P7F1H1 Seq. 30: P7F1H2	∅
Neutral arousal	Seq. 07: P7H4H4 Seq. 19: P8F4H1	Seq. 15: P4F4H1 Seq. 22: P2F4H4 Seq. 25: P3F4H4 Seq. 34: P3F4H1	Seq. 02: P3F4H2 Seq. 26: P4F1H2
High level of arousal	Seq. 01: P1F3H4 Seq. 05: P7F3H3 Seq. 06: P7F4H3 Seq. 09: P9F2H3 Seq. 10: P9F3H3 Seq. 11: P2F2H3 Seq. 13: P3F3H3 Seq. 14: P3F4H3 Seq. 16: P4F4H3 Seq. 17: P5F3H1 Seq. 18: P7F3H1 Seq. 20: P9F1H3 Seq. 28: P6F3H4 Seq. 31: P1F3H1 Seq. 32: P2F4H3 Seq. 33: P2F2H3 Seq. 35: P4F1H3 Seq. 37: P6F3H3 Seq. 39: P7F4H3 Seq. 40: P8F3H4	Seq. 23: P3F1H2 Seq. 24: P3F3H2 Seq. 38: P7F1H4	Seq. 03: P7F2H2 Seq. 04: P7F2H4 Seq. 08: P9F1H2 Seq. 12: P2F4H2 Seq. 21: P2F1H2 Seq. 36: P4F2H2

It also seems that when the avatar raises his eyebrows (sequences associated with the fourth facial expression, i.e., P*F4H*), the level of arousal is perceived as neutral, except if shaking his head at the same time (sequences associated with the fourth facial expression and the third head movement, i.e., P*F4H3). In this case, it is perceived as having a high level of arousal.

Some links between emotional valence and the level of arousal can be inferred. If the avatar's emotional valence is perceived as positive or negative, the associated level of arousal is never low.

This confirms the second hypothesis.

4.3 Comparison Between Low-end and High-end Headsets

This section investigates whether there is a difference in the confidence of assessing emotional valence and level of arousal when different headsets are used. T-tests were used to answer these questions. For consistency, only the results for the third set of non-verbal behaviours are compared. As explained in Section 3.3, both low-end and high-end headsets were used for the sequences of the third set. Therefore, only Sequences 21 to 40 will be considered in this section. The results of the comparison between these two headsets in terms of sense of presence are presented in Section 4.5).

The confidence level for the valence rating does not differ between the low-end headset ($mean_{low} = 5.972$) and high-end headset ($mean_{high} = 5.96$) (p-value=0.863). The confidence level for the arousal rating does not differ between the low-end headset ($mean_{low} = 6.252$) and the high-end headset ($mean_{high} = 6.266$) (p-value=0.820). For both tests, the headset used does not influence the confidence level, and confirms Hypothesis 3a.

4.4 Comparison Between Photo-realistic and Cartoon Avatars

Valence does not differ between the cartoon and the photo-realistic avatars ($mean_{cartoon} = 3.476$, $mean_{photo} = 3.519$, and p-value=0.427), nor does arousal ($mean_{cartoon} = 4.681$, $mean_{photo} = 4.687$, and p-value=0.427). For both tests, the non-verbal behaviours of the avatars are perceived in the same way for valence and arousal, regardless of the quality of the graphics used to represent them.

Finally, the seek was to determine whether the confidence level improves when assessing the level of valence and arousal for photo-realistic avatars. The confidence level for the valence rating is statistically different between the cartoon avatars and photo-realistic avatars ($mean_{cartoon} = 5.729$, $mean_{photo} = 6.228$, and p-value < 0.001), as is the confidence level for the arousal rating ($mean_{cartoon} = 6.067$, $mean_{photo} = 6.577$, and p-value < 0.001). Both tests confirmed Hypothesis 4.

4.5 Sense of Presence

Overall, participants rated presence in virtual reality as 66.76% on average based on the Gatineau Presence Questionnaire. Participants found their experience half realistic (average of 51.1%), and the virtual reality environment highly artificial (average of 84.16%). They forgot about their presence in the actual room (average of 58.32% for spatial awareness). The average of these four items is 65.085%.

T-tests, on the results of the Gatineau Presence Questionnaire for both the low-end and high-end headsets, were performed. On average, presence in virtual reality does not differ between headsets ($mean_{low} = 64.78\%$, $mean_{high} = 65.35\%$, and p-value $= 0.962$). Apart from the level of artificiality ($mean_{low} = 83.33\%$, $mean_{high} = 85.58\%$, and p-value $= 0.498$), using the high-end headset compared to the low-end headset increases presence in virtual reality ($mean_{low} = 62.46\%$, $mean_{high} = 73.14\%$, and p-value $= 0.002$), the level of realism ($mean_{low} = 46.44\%$, $mean_{high} = 57.3\%$, and p-value $= 0.018$), and decreases spatial awareness[5] ($mean_{low} = 66.9\%$, $mean_{high} = 45.4\%$, and p-value < 0.001). Thus, there is a significant difference for presence, realism, and spatial awareness. Only artificiality does not seem to have changed, which is confirmed by a t-test. As a result, using the high-end headset improves the sense of presence in general[6]. This supports Hypothesis 3b.

5 Qualitative Assessment

In this study, participants were given the opportunity to provide written comments on the avatars as they watched the sequences, allowing them to nuance their responses or provide valuable insights into their experiences. This section is not intended to be exhaustive, but rather to highlight certain aspects mentioned in the previous sections or add further insights.

First, participants felt able to imagine realistic situations that went beyond the sense of presence. As one participant noted, "This is an attitude that my boss could have."

Second, participants found it more difficult to evaluate emotions due to the absence of micro-expressions. One stated, "There are not enough micro facial expressions". On the other hand, the high-end headset provided a more detailed avatar, which interestingly made participants perplexed. As one participant mentioned, "The corner of the mouth changes everything. It looks like he is mocking or if we agree on a sensitive issue. Never sure about the level of alertness."

Third, some comments related to the potential difficulty of interpreting some expressions. Three different participants who watched an avatar smiling and shaking her head mentioned"The smile doesn't go with the head movement, it looks like mockery", "Her smile looks fake", and "I don't like her wry smile". Cultural specificities also played a role in participants' interpretations. Without context, it was sometimes difficult to determine the meaning behind certain behaviours. As one participant pointed out, "If I had a European person in front of me, they would say yes, but if it was an Indian person, they would say no", when referring to a shake of the head.

In addition, participants sometimes expressed their own emotional reactions to the avatar's emotions, which was not a specific aspect of the study. As one participant noted, "The sighs suggest an indifferent behaviour, which annoys me" or "He's scary, he doesn't look happy, I want to run away".

Finally, participants also expressed their own perceptions of the avatar's emotions: "She looks quite angry" or "He looks happy and joyful".

6 Discussion

This study examines how people perceive a virtual audience provided by virtual reality technology and how to select non-verbal behaviours to faithfully represent a range of audience reactions. The main contributions can be summarised as follows.

First, the study of perceived emotional valence and arousal for predefined parameters individually and in combination provides insights into how different combinations of non-verbal behaviours can be used to express specific levels of valence and arousal. Second, the comparison of the valence and arousal results when using low-end versus high-end virtual reality headsets shows that the sense of presence is improved with high-end headsets. Third, the comparison of the valence and arousal results when using cartoon avatars versus photo-realistic avatars shows that photo-realistic avatars improve the confidence level of participants' judgments without changing their assessment of valence and arousal. The results provide a typology of perceptions of valence and arousal, while

[5] This means that with the high-end headset, participants are more likely to forget they are in the actual room.

[6] In the Gatineau Presence Questionnaire, spatial awareness is rated in reverse order to presence in virtual reality and level of realism.

clearly identifying the sequences associated with positive, neutral, and negative valence, and a low, neutral, or high level of arousal.

This study builds on the results of Chollet and Scherer's (2017) work focused on participants creating combinations of non-verbal behaviours from a list (body postures, facial expressions, and head movements) to express the desired level of arousal and valence. In the present study, the focus is on immersed users' perceptions of some of the resulting non-verbal combinations in virtual reality rather than only on how creators might build such combinations in a two/three-dimensional setting (as in Chollet and Scherer (2017)). Moreover, their experiment was conducted on a flat screen via the web, rather than in a full virtual reality setting (with a head-mounted display), so it is unknown whether their results can be generalised to virtual reality. The aim of this present study is to determine whether the experiment medium influences the results. Another area of interest is the measurement of valence and arousal for specific non-verbal behaviours separately.

An interesting observation from these findings is the perception of smiles. In this study, smiles are primarily neutral, and raised eyebrows are associated with negative valence. This finding is notable because smiles are typically perceived as positive. However, based on participants' comments, this may be due to the fact that participants cannot distinguish between genuine and fake smiles. This ambiguity may be influenced by the combination of the smile with other non-verbal behaviours in the sequence, such as head shaking. This analysis is consistent with Krumhuber et al.'s (2009) work on smiles.

Furthermore, this study shows that photo-realistic avatars can increase confidence levels by making it easier for participants to associate the avatar's non-verbal behaviours with real-life experiences (as also highlighted in the qualitative assessment and in De Keyser et al., 2019). This idea is consistent with Seymour et al. (2021) who find increased trustworthiness and affinity with human-realistic avatars. Despite the uncanny valley phenomenon implying an aversion to near-realistic avatars, our findings are consistent with Seymour et al. (2021). Indeed, they conclude that it is now possible to overcome the uncanny valley using real-time-rendered human realistic avatars, and we believe our similar results support this notion.

In addition, this study shows that emotional valence is easily perceived regardless of the headset used, although high-end headsets provide more accurate details. Valence may not be directly affected when non-verbal behaviour combinations are easily visible. However, arousal levels can be influenced by facial expressions, with high-end headsets providing clearer details, particularly around the eyes, resulting in higher perceived arousal. This is consistent with Orús et al. (2021), as high-end headsets enhance the participant's sense of presence, making it easier to associate the avatar's non-verbal behaviours with real-life experiences. Surprisingly, high-end headsets do not increase confidence levels, possibly because their effect is already evident in the arousal assessment. Another explanation can be linked with the uncanney valley effect, as explained in Di Natale et al. (2023). The comparison between low-end and high-end headsets highlights the effectiveness of future virtual reality training environments using simple Cardboard, addressing issues of affordability and mass diffusion.

In this study, the Gatineau Presence Questionnaire was used to measure the sense of presence in virtual reality. The average score obtained is 65.085%, consistent with the studies of Laforest et al. (2016) and Bouchard et al. (2017) who report scores between 48.11% and 65.73%. The lower than expected presence score in this study may be due to the time and disorientation of leaving and re-entering the virtual reality environment, as explained in Schwind et al. (2019). Moreover, asking participants about their sense of presence can lead to biased results and the so-called break-in-presence, as discussed in Jerald (2015) and Slater and Steed (2000).

Overall, this study demonstrates the complexity of interpreting non-verbal cues and emotions in virtual environments. The participants' comments shed light on the subjective experiences and the impact of various factors on their perceptions.

These findings have significant implications for the use of virtual reality technology in training frontline staff and improving their interactions with consumers. By identifying specific non-verbal behaviours associated with certain levels of emotional valence and arousal, this study offers guidance on designing avatars in virtual reality for training purposes. For example, using photo-realistic avatars can improve participants' confidence in their valence and arousal judgements, potentially leading to more effective training outcomes. Similarly, using high-end virtual reality headsets can enhance the sense of presence, resulting in a more realistic training experience. Understanding the link between non-verbal behaviours and emotional valence and arousal can help tailor training content to specific situations and interactions that employees may encounter with customers. This can lead to more targeted and effective training programs, ultimately resulting in improved employee behaviours and better customer experiences.

7 Limitations and Future Research

This study has a number of limitations that should be taken into account when interpreting the results. One such limitation is the fact that certain sequences were performed exclusively by either female or male avatars, despite attempts to control for gender-related factors. This limitation may have influenced the results, and future research could address this issue by using a more balanced design that includes an equal number of male and female avatars executing all sequences. In addition, future research could explore the complex ways in which gender may interact with other factors, such as emotional valence and arousal levels, to better understand the impact of gender on avatar non-verbal behaviours. Another limitation relates to the need for caution when interpreting the results and their generalizability to different populations. For instance, the interpretation of nods varies across cultures, and this factor may have influenced the present results. Moreover, the number of combinations of animations that we were able to test in this study was limited due to the nature of the experiment based on true virtual reality immersion. Finally, the methodology is limited by the need for participants to leave and re-enter the virtual reality environment to complete the questionnaire, which may also have influenced the results.

The participants' comments also highlight the complexity of interpreting emotions in virtual environments and the need to consider cultural influences when analysing virtual interactions. This unexpected finding provides insights into how participants perceive and internalise emotions, offering potential avenues for future research.

In conclusion, this study provides valuable insights into the emotional design of avatar non-verbal behaviours in the virtual reality context. Furthermore, thanks to the typology developed in this paper, it is now clear which sequences to choose in order to demonstrate specific levels of arousal (low, neutral, and high) and emotional valence (negative, neutral, and positive). This can be very useful when designing new virtual reality environments for training to determine the non-verbal behaviour to employ in response to a particular situation. As well as contributing to understanding virtual interactions, this research opens up new possibilities for applications, such as public speaking training, highlighting the virtual reality potential to simulate future scenarios.

Future work will therefore aim to use these avatar non-verbal behaviours to elicit appropriate responses from the avatars during a user's training in virtual reality, such as gradually increasing the induced anxiety or rewarding good performance. There will be two types of avatars: other participants (trainees, experts, teachers, etc.) and artificial intelligence driven avatars (Butt et al., 2021). A study on automatic methods based on statistical, machine learning, and natural language processing methods to implement real-time feedback from the audience to the speaker's presentation is under consideration. While the first application being considered next is the training in front of multiple avatars (public speaking in general), the present results could be useful in many other contexts.

References

Amin, A., Gromala, D., Tong, X., and Shaw, C. (2016). Immersion in cardboard VR compared to a traditional head-mounted display. In *International Conference on Virtual, Augmented and Mixed Reality*, pages 269–276. Springer.

Argyle, M. and Ingham, R. (1972). Gaze, mutual gaze, and proximity. *Semiotica*, 6(1):32–49.

Babin, B. J. and Attaway, J. S. (2000). Atmospheric affect as a tool for creating value and gaining share of customer. *Journal of Business Research*, 49(2):91–99.

Batrinca, L., Stratou, G., Shapiro, A., Morency, L.-P., and Scherer, S. (2013). Cicero-towards a multimodal virtual audience platform for public speaking training. In *International Workshop on Intelligent Virtual Agents*, pages 116–128. Springer.

Beck, J., Rainoldi, M., and Egger, R. (2019). Virtual reality in tourism: A state-of-the-art review. *Tourism Review*, 74(3):586–612.

Belanche, D., Casaló, L. V., Schepers, J., and Flavián, C. (2021). Examining the effects of robots' physical appearance, warmth, and competence in frontline services: The humanness-value-loyalty model. *Psychology & Marketing*, 38(12):2357–2376.

Bonetti, F., Warnaby, G., and Quinn, L. (2018). Augmented reality and virtual reality in physical and online retailing: A review, synthesis and research agenda. *Augmented Reality and Virtual Reality*, pages 119–132.

Bouchard, S., Dumoulin, S., Robillard, G., Guitard, T., Klinger, E., Forget, H., Loranger, C., and Roucaut, F.-X. (2017). Virtual reality compared with in vivo exposure in the treatment of social anxiety disorder: A three-arm randomised controlled trial. *The British Journal of Psychiatry*, 210(4):276–283.

Bouchard, S. and Robillard, G. (2019). Validation canadienne-française du Gatineau Presence Questionnaire auprès d'adultes immergés en réalité virtuelle. *87e Congrès de l'ACFAS, Québec, mai 2019.*

Boyd, D. E. and Koles, B. (2019). An introduction to the special issue "virtual reality in marketing": Definition, theory and practice. *Journal of Business Research*, 100:441–444.

Brady, M. K., Robertson, C. J., and Cronin, J. J. (2001). Managing behavioral intentions in diverse cultural environments: An investigation of service quality, service value, and satisfaction for american and ecuadorian fast-food customers. *Journal of International Management*, 7(2):129–149.

Butt, A. H., Ahmad, H., Goraya, M. A., Akram, M. S., and Shafique, M. N. (2021). Let's play: Me and my ai-powered avatar as one team. *Psychology & Marketing*, 38(6):1014–1025.

Cafaro, A., Vilhjálmsson, H. H., and Bickmore, T. (2016). First impressions in human–agent virtual encounters. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 23(4):1–40.

Chollet, M., Marsella, S., and Scherer, S. (2021). Training public speaking with virtual social interactions: Effectiveness of real-time feedback and delayed feedback. *Journal on Multimodal User Interfaces*, 16:17–29.

Chollet, M., Massachi, T., and Scherer, S. (2016a). Investigating the physiological responses to virtual audience behavioral changes a stress-aware audience for public speaking training. In *IVA 2017 Workshop on Physiologically-Aware Virtual Agents.*

Chollet, M. and Scherer, S. (2017). Perception of virtual audiences. *IEEE Computer Graphics and Applications*, 37(4):50–59.

Chollet, M., Wörtwein, T., Morency, L.-P., and Scherer, S. (2016b). A multimodal corpus for the assessment of public speaking ability and anxiety. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 488–495.

Chollet, M., Wörtwein, T., Morency, L.-P., Shapiro, A., and Scherer, S. (2015). Exploring feedback strategies to improve public speaking: An interactive virtual audience framework. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 1143–1154.

Cowan, K. and Ketron, S. (2019). A dual model of product involvement for effective virtual reality: The roles of imagination, co-creation, telepresence, and interactivity. *Journal of Business Research*, 100:483–492.

Craske, M. G., Treanor, M., Conway, C. C., Zbozinek, T., and Vervliet, B. (2014). Maximizing exposure therapy: An inhibitory learning approach. *Behaviour Research and Therapy*, 58:10–23.

De Keyser, A., Köcher, S., Alkire, L., Verbeeck, C., and Kandampully, J. (2019). Frontline service technology infusion: Conceptual archetypes and future research directions. *Journal of Service Management*, 30(1):156–183.

Deng, X., Unnava, H. R., and Lee, H. (2019). "too true to be good?" when virtual reality decreases interest in actual reality. *Journal of Business Research*, 100:561–570.

Di Natale, A. F., Simonetti, M. E., La Rocca, S., and Bricolo, E. (2023). Uncanny valley effect: A qualitative synthesis of empirical research to assess the suitability of using virtual faces in psychological research. *Computers in Human Behavior Reports*, page 100288.

Dwivedi, Y. K., Hughes, L., Wang, Y., Alalwan, A. A., Ahn, S. J., Balakrishnan, J., Barta, S., Belk, R., Buhalis, D., Dutot, V., et al. (2023). Metaverse marketing: How the metaverse will shape the future of consumer research and practice. *Psychology & Marketing*, 40(4):750–776.

Eckert, D. and Mower, A. (2020). The effectiveness of virtual reality soft skills training in the enterprise: A study. *PwC Public Report*.

Farrell, A. M., Souchon, A. L., and Durden, G. R. (2001). Service encounter conceptualisation: Employees' service behaviours and customers' service quality perceptions. *Journal of Marketing Management*, 17(5-6):577–593.

Flavián, C., Ibáñez-Sánchez, S., and Orús, C. (2019). The impact of virtual, augmented and mixed reality technologies on the customer experience. *Journal of business research*, 100:547–560.

Fraune, M. R., Oisted, B. C., Sembrowski, C. E., Gates, K. A., Krupp, M. M., and Šabanović, S. (2020). Effects of robot-human versus robot-robot behavior and entitativity on anthropomorphism and willingness to interact. *Computers in Human Behavior*, 105:106220.

Freeman, D., Reeve, S., Robinson, A., Ehlers, A., Clark, D., Spanlang, B., and Slater, M. (2017). Virtual reality in the assessment, understanding, and treatment of mental health disorders. *Psychological Medicine*, 47(14):2393–2400.

Glémarec, Y., Bosser, A.-G., Buche, C., Lugrin, J.-L., Landeck, M., Latoschik, M., and Chollet, M. (2019). A scalability benchmark for a virtual audience perception model in virtual reality. In *25th ACM Symposium on Virtual Reality Software and Technology*, pages 1–1.

Goberman, A. M., Hughes, S., and Haydock, T. (2011). Acoustic characteristics of public speaking: Anxiety and practice effects. *Speech Communication*, 53(6):867–876.

Hainey, T., Connolly, T. M., Boyle, E. A., Wilson, A., and Razak, A. (2016). A systematic literature review of games-based learning empirical evidence in primary education. *Computers & Education*, 102:202–223.

Hennig-Thurau, T., Aliman, D. N., Herting, A. M., Cziehso, G. P., Linder, M., and Kübler, R. V. (2022). Social interactions in the metaverse: Framework, initial evidence, and research roadmap. *Journal of the Academy of Marketing Science*, pages 1–25.

Hilken, T., Chylinski, M., de Ruyter, K., Heller, J., and Keeling, D. I. (2022). Exploring the frontiers in reality-enhanced service communication: From augmented and virtual reality to neuro-enhanced reality. *Journal of Service Management*, 33(4/5):657–674.

Howard, M. C. (2017). A meta-analysis and systematic literature review of virtual reality rehabilitation programs. *Computers in Human Behavior*, 70:317–327.

Howard, M. C., Gutworth, M. B., and Jacobs, R. R. (2021). A meta-analysis of virtual reality training programs. *Computers in Human Behavior*, 121:106808.

Jerald, J. (2015). *The VR book: Human-centered design for virtual reality*. Morgan & Claypool.

Jingen Liang, L. and Elliot, S. (2021). A systematic review of augmented reality tourism research: What is now and what is next? *Tourism and Hospitality Research*, 21(1):15–30.

Kang, H. and Kim, H. K. (2020). My avatar and the affirmed self: Psychological and persuasive implications of avatar customization. *Computers in Human Behavior*, 112:106446.

Kang, N., Brinkman, W.-P., van Riemsdijk, M., and Neerincx, M. (2016). The design of virtual audiences: Noticeable and recognizable behavioral styles. *Computers in Human Behavior*, 55:680–694.

Khandelwal, K. and Upadhyay, A. K. (2021). Virtual reality interventions in developing and managing human resources. *Human Resource Development International*, 24(2):219–233.

Kim, D. Y., Lee, H. K., and Chung, K. (2023). Avatar-mediated experience in the metaverse: The impact of avatar realism on user-avatar relationship. *Journal of Retailing and Consumer Services*, 73:103382.

Kim, S. Y., Schmitt, B. H., and Thalmann, N. M. (2019). Eliza in the uncanny valley: Anthropomorphizing consumer robots increases their perceived warmth but decreases liking. *Marketing Letters*, 30(1):1–12.

Kleinsmith, A., Rivera-Gutierrez, D., Finney, G., Cendan, J., and Lok, B. (2015). Understanding empathy training with virtual patients. *Computers in Human Behavior*, 52:151–158.

Knapp, M. L., Hall, J. A., and Horgan, T. G. (2013). *Nonverbal communication in human interaction*. Cengage Learning.

Koles, B. and Nagy, P. (2021). Digital object attachment. *Current Opinion in Psychology*, 39:60–65.

Krumhuber, E., Manstead, A. S., Cosker, D., Marshall, D., and Rosin, P. L. (2009). Effects of dynamic attributes of smiles in human and synthetic faces: A simulated job interview setting. *Journal of Nonverbal Behavior*, 33(1):1–15.

Laforest, M., Bouchard, S., Crétu, A.-M., and Mesly, O. (2016). Inducing an anxiety response using a contaminated Virtual environment: Validation of a therapeutic tool for obsessive–compulsive disorder. *Frontiers in ICT*, 3:18.

Larivière, B., Bowen, D., Andreassen, T. W., Kunz, W., Sirianni, N. J., Voss, C., Wünderlich, N. V., and De Keyser, A. (2017). "service encounter 2.0": An investigation into the roles of technology, employees and customers. *Journal of Business Research*, 79:238–246.

Letheren, K., Jetten, J., Roberts, J., and Donovan, J. (2021). Robots should be seen and not heard... sometimes: Anthropomorphism and ai service robot interactions. *Psychology & Marketing*, 38(12):2393–2406.

Leveau, P.-H. and Camus, S. (2023). Embodiment, immersion, and enjoyment in virtual reality marketing experiences. *Psychology & Marketing*, 40(7):1329–1343.

Liao, H. and Chuang, A. (2004). A multilevel investigation of factors influencing employee service performance and customer outcomes. *Academy of Management Journal*, 47(1):41–58.

Lloyd, A. E. and Luk, S. T. (2011). Interaction behaviors leading to comfort in the service encounter. *Journal of Services marketing*, 25(3):176–189.

Loureiro, S. M. C., Guerreiro, J., Eloy, S., Langaro, D., and Panchapakesan, P. (2019). Understanding the use of virtual reality in marketing: A text mining-based review. *Journal of Business Research*, 100:514–530.

Lugrin, J.-L., Latoschik, M. E., Habel, M., Roth, D., Seufert, C., and Grafe, S. (2016). Breaking bad behaviors: A new tool for learning classroom management using virtual reality. *Frontiers in ICT*, 3:26.

Maria Stock, R., Jong, A. d., and Zacharias, N. A. (2017). Frontline employees' innovative service behavior as key to customer loyalty: Insights into fles' resource gain spiral. *Journal of Product Innovation Management*, 34(2):223–245.

Miao, F., Kozlenkova, I. V., Wang, H., Xie, T., and Palmatier, R. W. (2022). An emerging theory of avatar marketing. *Journal of Marketing*, 86(1):67–90.

Murphy, J., Gretzel, U., and Pesonen, J. (2019). Marketing robot services in hospitality and tourism: The role of anthropomorphism. *Journal of Travel & Tourism Marketing*, 36(7):784–795.

Nash, E. B., Edwards, G. W., Thompson, J. A., and Barfield, W. (2000). A review of presence and performance in virtual environments. *International Journal of Human-Computer Interaction*, 12(1):1–41.

Orús, C., Ibáñez-Sánchez, S., and Flavián, C. (2021). Enhancing the customer experience with virtual and augmented reality: The impact of content and device type. *International Journal of Hospitality Management*, 98:103019.

Pertaub, D.-P., Slater, M., and Barker, C. (2002). An experiment on public speaking anxiety in response to three different types of virtual audience. *Presence*, 11(1):68–78.

Poeschl, S. (2017). Virtual reality training for public speaking — A QUEST-VR framework validation. *Frontiers in ICT*, 4:13.

Rafaeli, A., Altman, D., Gremler, D. D., Huang, M.-H., Grewal, D., Iyer, B., Parasuraman, A., and de Ruyter, K. (2017). The future of frontline research: Invited commentaries. *Journal of Service Research*, 20(1):91–99.

Remacle, A., Bouchard, S., Etienne, A.-M., Rivard, M.-C., and Morsomme, D. (2021). A virtual classroom can elicit teachers' speech characteristics: Evidence from acoustic measurements during in vivo and in virtuo lessons, compared to a free speech control situation. *Virtual Reality*, 25:935–944.

Remacle, A.-F., Bouchard, S., and Morsomme, D. (2023). Can teaching simulations in a virtual classroom help trainee teachers to develop oral communication skills and self-efficacy? a randomized controlled trial. *Computers Education*, 200:104810.

Robinson, S., Orsingher, C., Alkire, L., De Keyser, A., Giebelhausen, M., Papamichail, K. N., Shams, P., and Temerak, M. S. (2020). Frontline encounters of the ai kind: An evolved service encounter framework. *Journal of Business Research*, 116:366–376.

Scheveneels, S., Boddez, Y., Van Daele, T., and Hermans, D. (2019). Virtually unexpected: No role for expectancy violation in virtual reality exposure for public speaking anxiety. *Frontiers in Psychology*, 10:2849.

Schrader, C. (2019). Creating avatars for technology usage: Context matters. *Computers in Human Behavior*, 93:219–225.

Schwind, V., Knierim, P., Haas, N., and Henze, N. (2019). Using presence questionnaires in virtual reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–12.

Selzer, M. N., Gazcon, N. F., and Larrea, M. L. (2019). Effects of virtual presence and learning outcome using low-end virtual reality systems. *Displays*, 59:9–15.

Seufert, C., Oberdörfer, S., Roth, A., Grafe, S., Lugrin, J.-L., and Latoschik, M. E. (2022). Classroom management competency enhancement for student teachers using a fully immersive virtual classroom. *Computers & Education*, 179:104410.

Seymour, M., Riemer, K., and Kay, J. (2018). Actors, avatars and agents: Potentials and implications of natural face technology for the creation of realistic visual presence. *Journal of the Association for Information Systems*, 19(10):4.

Seymour, M., Yuan, L. I., Dennis, A., Riemer, K., et al. (2021). Have we crossed the uncanny valley? understanding affinity, trustworthiness, and preference for realistic digital humans in immersive environments. *Journal of the Association for Information Systems*, 22(3):9.

Sibilla, F. and Mancini, T. (2018). I am (not) my avatar: A review of the user-avatar relationships in massively multiplayer online worlds. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 12(3):article 4.

Slater, M. (2003). A note on presence terminology. *Presence Connect*, 3(3):1–5.

Slater, M. and Steed, A. (2000). A virtual presence counter. *Presence*, 9(5):413–434.

Smith, S. and Wheeler, J. (2002). *Managing the customer experience: Turning customers into advocates*. Pearson Education.

Solomon, M. R., Surprenant, C., Czepiel, J. A., and Gutman, E. G. (1985). A role theory perspective on dyadic interactions: The service encounter. *Journal of Marketing*, 49(1):99–111.

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

van Bennekom, M. J., de Koning, P. P., and Denys, D. (2017). Virtual reality objectifies the diagnosis of psychiatric disorders: A literature review. *Frontiers in Psychiatry*, 8:163.

Van Pinxteren, M. M., Wetzels, R. W., Rüger, J., Pluymaekers, M., and Wetzels, M. (2019). Trust in humanoid robots: Implications for services marketing. *Journal of Services Marketing*, 33(4):507–518.

Vlachopoulos, D. and Makri, A. (2017). The effect of games and simulations on higher education: A systematic literature review. *International Journal of Educational Technology in Higher Education*, 14(1):1–33.

Wang, W. and Benbasat, I. (2005). Trust in and adoption of online recommendation agents. *Journal of the Association for Information Systems*, 6(3):4.

Weech, S., Kenny, S., and Barnett-Cowan, M. (2019). Presence and cybersickness in virtual reality are negatively related: A review. *Frontiers in Psychology*, 10:158.

Wörtwein, T., Chollet, M., Schauerte, B., Morency, L.-P., Stiefelhagen, R., and Scherer, S. (2015). Multimodal public speaking performance assessment. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, pages 43–50.

Yao, Q., Kuai, L., and Jiang, L. (2023). Effects of the anthropomorphic image of intelligent customer service avatars on consumers' willingness to interact after service failures. *Journal of Research in Interactive Marketing*, ahead-of-print.

A Appendix

A.1 Analysis Of Postures, Facial Expressions, And Head Movements

A Chi-square test was conducted to see if the participants' responses were evenly distributed between negative interpretations (answers 1 to 3 in the Likert-type scale) and positive interpretations (answers 5 to 7 in Likert-type scale)[7]; first for valence (Table A1) and then for the arousal (Table A2). The aim is to determine whether the difference between the proportions of positive (denoted with p_+) and negative interpretations (denoted with p_-) are statistically significant. The p-values obtained for this test are presented in the two tables. If the difference between the proportions of positive and negative interpretations is not statistically significant, it is denoted with "/". If not, and if there are more positive than negative interpretations, it is denoted with "+". If there were more negative than positive interpretations, it is denoted with "-". The two percentages correspond to the proportion of participants with a negative interpretation (1, 2, or 3) and a positive interpretation (5, 6, or 7). The proportion of neutral answers (4) is thus not in the table but can be easily derived[8].

Table A1: P-values and interpretations for valence per non-verbal behaviour. (Please refer to Table 1 and Section 4.1 for the notations)

		Proportions		P-value	Interpretation
		p_-	p_+		
Postures					
P1	Backward posture – Arms crossed	48%	32%	< 0.001	-
P2	Backward posture – Arms down	49%	30%	< 0.001	-
P3	Backward posture – Arms behind the head	55%	15%	< 0.001	-
P4	Upright posture – Hand on hand	40%	39%	0.887	/
P5	Upright posture – Hands together	56%	22%	< 0.001	-
P6	Upright posture – Hands separated in front	64%	21%	< 0.001	-
P7	Forward posture – Hands together	38%	38%	0.918	/
P8	Forward posture – Arms down	52%	21%	< 0.001	-
P9	Forward posture – Arms crossed	50%	30%	0.002	-
Facial expressions					
F1	None	36%	41%	0.185	/
F2	Smiling	38%	43%	0.305	/
F3	Frowning	57%	26%	< 0.001	-
F4	Eyebrows raised	50%	23%	< 0.001	-
Head movements					
H1	None	50%	28%	< 0.001	-
H2	Nod	22%	60%	< 0.001	+
H3	Shake	70%	14%	< 0.001	-
H4	Questioning	38%	29%	0.028	-

[7] Throughout this paper, and for clarity, all responses from 1 to 3 in the Likert-type scale will be referred to as "negative" and those from 5 to 7 in the Likert-type scale as "positive". The words "negative" and "positive" must be used with caution. For valence, it can be understood as having negative and positive levels. For arousal, it can be understood as low and high levels.

[8] 100 % minus the two other percentages.

Table A2: P-values and interpretations of the test where the null hypothesis is $p_- = p_+$ for arousal per non-verbal behaviour. (Please refer to Table 1 and Section 4.1 for the notations)

		Proportions p_-	p_+	P-value	Interpretation
Postures					
P1	Backward posture – Arms crossed	21%	52%	0.0006	+
P2	Backward posture – Arms down	18%	68%	< 0.001	+
P3	Backward posture – Arms behind the head	26%	51%	< 0.001	+
P4	Upright posture – Hand on hand	23%	60%	< 0.001	+
P5	Upright posture – Hands together	43%	40%	0.8262	/
P6	Upright posture – Hands separated in front	16%	70%	< 0.001	+
P7	Forward posture – Hands together	29%	56%	< 0.001	+
P8	Forward posture – Arms down	34%	53%	0.0536	/
P9	Forward posture – Arms crossed	24%	60%	< 0.001	+
Facial expressions					
F1	None	31%	52%	< 0.001	+
F2	Smiling	18%	66%	< 0.001	+
F3	Frowning	22%	58%	< 0.001	+
F4	Eyebrows raised	29%	55%	< 0.001	+
Head movements					
H1	None	33%	50%	0.0004	+
H2	Nod	22%	60%	< 0.001	+
H3	Shake	18%	65%	< 0.001	+
H4	Questioning	35%	47%	0.0049	+