



HAL
open science

Impact of the Nonverbal Behavior of Virtual Audience on Users' Perception of Social Attitudes

Marion Ristorcelli, Alexandre d'Ambra, Jean-Marie Pergandi, Rémy Casanova, Magalie Ochs

► **To cite this version:**

Marion Ristorcelli, Alexandre d'Ambra, Jean-Marie Pergandi, Rémy Casanova, Magalie Ochs. Impact of the Nonverbal Behavior of Virtual Audience on Users' Perception of Social Attitudes. 2024 International Conference on Advanced Visual Interfaces (AVI 2024), Jun 2024, Arenzano, Italy. pp.16, 10.1145/3656650.3656687 . hal-04613261

HAL Id: hal-04613261

<https://hal.science/hal-04613261>

Submitted on 18 Jun 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

Impact of the Nonverbal Behavior of Virtual Audience on Users' Perception of Social Attitudes

Marion RISTORCELLI
marion.ristorcelli@lis-lab.fr
Aix Marseille Univ, CNRS, LIS UMR
7020, France

Alexandre D'AMBRA
alexandre.D-AMBRA@univ-amu.fr
Aix Marseille Univ, CNRS, ISM UMR
7287, France

Jean-Marie PERGANDI
jean-marie.pergandi@univ-amu.fr
Aix Marseille Univ, CNRS, ISM UMR
7287, France

Rémy CASANOVA
remy.casanova@univ-amu.fr
Aix Marseille Univ, CNRS, ISM UMR
7287, France

Magalie OCHS
magalie.ochs@lis-lab.fr
Aix Marseille Univ, CNRS, LIS UMR
7020, France

ABSTRACT

In a virtual reality public speaking training system, it is essential to control the audience's nonverbal behavior in order to simulate different attitudes. The virtual audience's social attitude is generally represented by a two-dimensional valence-arousal model describing the opinion and engagement of virtual characters. In this article, we argue that the valence-arousal representation is not sufficient to describe the user's perception of a virtual character's social attitude. We propose a three-dimensional model by dividing the valence axis into two dimensions representing the epistemic and affective stance of the virtual character, reflecting the character's agreement and emotional reaction. To assess the perception of the virtual characters' nonverbal behavior on these two new dimensions, we conducted a perceptive study in virtual reality with 44 participants who evaluated 50 animations combining multimodal nonverbal behavioral signals such as head movements, facial expressions, gaze direction and body posture. The results of our experiment show that, in fact, the valence axis should be divided into two axes to take into account the perception of the virtual character's epistemic and affective stance. Furthermore, the results show that one behavioral signal is predominant for the evaluation of each dimension: head movements for the epistemic dimension and facial expressions for the affective dimension. These results provide useful guidelines for designing the nonverbal behavior of a virtual audience for social attitudes' simulation.

CCS CONCEPTS

• **Human-centered computing** → **Virtual reality**.

KEYWORDS

Virtual reality, Virtual character, Social attitude, Nonverbal behavioral signals, Public speaking

ACM Reference Format:

Marion RISTORCELLI, Alexandre D'AMBRA, Jean-Marie PERGANDI, Rémy CASANOVA, and Magalie OCHS. 2024. Impact of the Nonverbal Behavior of Virtual Audience on Users' Perception of Social Attitudes. In *Proceedings of International Conference on Advanced Visual Interfaces (AVI '24)*. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3656650.3656687>

1 INTRODUCTION

Nowadays, public speaking is a primordial social skill. It appears in different context in our daily life both professionally and personally [23]. Whether at a job interview, looking for an internship or giving an oral presentation to a medium or large audience, speakers need to follow certain rules to make a good presentation [1], but they also have to adapt to the audience's behavior in order to maintain their attention and interest [8, 26, 27]. These competencies are not innate. We are not naturally good speakers and confident during a speech. This exercise requires a lot of preparation and training. However, we have relatively few tools for rehearsing public speaking in a context close to reality [11].

For some years now, Virtual Reality (VR) has emerged as a key tool for social skills training, and more particularly for public speaking training [6, 17, 18, 23, 31]. These tools are public speaking simulators. They integrate virtual environments, inspired by real-life situations (classroom, conference room), and are composed of virtual characters who play the role of spectators [5, 16]. This kind of tool is also widely used in therapy, particularly to treat glossophobia, which refers to a strong fear of public speaking. In this case, participants had to give several speeches successively in order to reduce anxiety via repeated, controlled exposure to the task as part of the Virtual Reality Exposure Therapy (VRET) [12, 33].

In virtual reality public speaking training system, the virtual audience is particularly important. The behavioral believability of the virtual audience is a key factor for the training [10]. Indeed, the virtual audience behavior may convey different social attitudes, such as interest or boredom [16], through their behavior (e.g. facial expressions, head movements, posture and gaze direction) and these simulated social attitudes have a strong impact on the user's perceived difficulty in speaking in public [28, 30]. Understanding how the social attitudes are perceived according to the virtual audience's behavior is therefore of key importance for the design of VR public speaking training system. Today, we use the valence-arousal

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

AVI '24, June 03–07, 2024, Arezano, Italy

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM

<https://doi.org/10.1145/3656650.3656687>

model to characterize the nonverbal behavior of a virtual character in terms of opinion (positive, neutral or negative valence) and engagement (low, neutral or high arousal) [5]. The social attitudes of a virtual character are therefore obtained from a combination of valence and arousal values [16]. However, in this work, we hypothesize that this model is not sufficient and can be improved to describe the user's perception of the social attitude of a virtual character.

In this article, our objective is to investigate how the virtual audience is perceived in terms of social attitudes by varying its nonverbal behavior. To go beyond the state of the art, we examine the perception of combinations of nonverbal behavioral signals, exploring new dimensions, such as the epistemic and affective stance, which could correspond to a subdivision of the valence dimension in the well-known valence-arousal model. For this purpose, we conducted a VR experiment in which 44 participants were asked to rate 12 statements using a 7-point Likert scale for 50 animations corresponding to different combinations of head movements, facial expressions, gaze direction and posture (arm and torso position).

This paper is organized as follows. In the next section, we present the work related to behavioral models of the virtual audience and the perception of the audience's nonverbal behaviors. In Section 3, we detail the experimental protocol and the materials used in our study. We present the results of the experiment Section 4 and discuss them Section 5. We conclude Section 6.

2 RELATED WORKS

2.1 Virtual Audience Behavioral Model

In VR public speaking training system domain, some researchers have investigated which behaviors are used to express different social attitudes [5, 13, 16, 21]. It is particularly important to know how these behaviors are perceived because they can have a strong impact on the user experience. For instance, an audience conveying a negative attitude, such as yawning, turning the head, leaning back and avoiding eye contact generates greater anxiety in the speaker than a positive virtual audience [28]. On the contrary, a virtual audience providing positive feedbacks through the virtual characters' nonverbal behavior induces more positive emotions and allows speakers to feel more comfortable during their presentation than in front of a negative virtual audience [30].

Several research works have investigated the relation between the audience's behaviors and the perceived attitudes. A seminal work has been conducted by Kang et al. [21]. They have recorded the behavior of real audiences in order to identify the associated perceived social attitude, such as boredom. Their study enabled them to highlight two dimensions that characterize a social attitude for a virtual audience: the valence and the arousal. The notion of valence refers to the audience's opinion, i.e. the positive or negative feelings they have about the speaker or the presentation [5]. The notion of arousal refers to the virtual characters' engagement in the presentation [16]. This representation has subsequently been used many times to model the nonverbal behavior of a virtual audience. Two major studies [5, 16] have shown that nonverbal signals, such as head movements, facial expressions, gaze direction and posture, can be associated to the perception of valence (negative, neutral, positive) and arousal (low, medium, high) [5], and that

an attitude can be represented precisely by a combination of the valence-arousal pair [15]. More recently, Etienne et al. [13] have identified precisely the combinations of nonverbal signals that should be used by a virtual character to convey a positive valence with a high level of arousal, or a negative valence with a neutral level of arousal.

Although the valence-arousal model seems to have made a place in the field of simulating the virtual audience's social attitude, it has certain limitations. Indeed, the valence dimension only focuses on the affective reaction to the speaker or the presentation, without taking into account the audience's agreement concerning what the speaker is saying. To consider these two aspects, the valence dimension should be divided into two sub-dimensions to represent both the affective reaction of the virtual audience and its agreement with the presentation. In the scientific literature on the face-to-face communication, researchers such as Chindamo et al. [4], distinguish between two types of stance: epistemic and affective. The epistemic stance corresponds to the degree of certainty expressed by the speaker regarding his/her speech, while the affective stance concerns the emotional state expressed by the interlocutor in response to his/her statement [4]. The same principle could therefore be used in the context of virtual audience to display its agreement and emotional state. In our case, the epistemic stance could be seen as the judgement or evaluation of the speaker's statement, which would itself be independent of the emotional reaction expressed by the interlocutor. It is possible, for example, to imagine a member of the audience disagreeing with the content of the presentation without necessarily expressing negative emotion. As described in the next section, the social attitude of the virtual audience is conveyed by a specific combination of nonverbal signals expressed by each virtual character.

2.2 Perception of Audience's Nonverbal Behavior

Some research concerning VR audiences has studied the perception of nonverbal behavior (facial expressions, head movements, gaze behavior, gestures and postures) in general, and some has focused more specifically on their perception in terms of opinion of the speaker and engagement in the presentation [5, 13, 16]. There is a consensus for certain signals, such as head movements, but we sometimes observe discrepancies, particularly regarding posture. In the following, we describe in greater detail the various research results concerning the perception of the nonverbal behavior of a virtual audience.

2.2.1 Head movements. According to Chollet et al. [5], the head movements are the most important signal to convey the audience valence and this information has been repeated and confirmed by other researchers [13, 16]. More precisely, a head nod conveys the positive valence of a virtual character in terms of opinion, whereas a head shake expresses a negative valence. In the literature, we also note that tilting the head is associated with a negative epistemic stance, showing the speaker's uncertainty [14]. In human-human interactions, it has also led to negative evaluations of several personality traits of a person, such as happiness [25]. In VR, head tilt is also associated with negative valence [13].

2.2.2 Facial expressions. Facial expressions are a strong signal for conveying audience valence [5]. Like the head nod, the smile is also associated with a positive valence. By contrast, frown, like head shake, is associated to a negative valence. In general, raised eyebrows reflect doubt and uncertainty [3], although this has not been verified in other major studies [5, 16].

2.2.3 Postures and Gestures. The posture characterizes the torso posture and the arms position.

Concerning the perception of the arms position, things are not clear-cut. It seems that arm-crossing is a behavior considered to be anti-social [34]. It's a sign of disagreement, dominance and disinterest from the audience [9, 26, 32]. In the study conducted by Chollet et al. [5], the authors associate this arm position with a closed posture and report that the posture contributes significantly to the valence assessment. However, this result is not confirmed by the study led by Glemarec et al. [16]. Closed gestures can be defined as gestures in which the hands are closely positioned. Conversely, open gestures are defined by hands far away from each other [35]. Similarly, other authors indicate that clasped or clenched hands also represent a closed posture [7]. Although the notion of open posture appears vague, all agree that crossed arms correspond to a closed posture and are a sign of disengagement [5, 35].

To go beyond the state of the art presented above, in the work presented in this paper, we aim at exploring two research questions: (1) On the one hand, we investigate the two-dimensional representation of audience's social attitudes. We argue that the valence-arousal representation is not sufficient to describe the different social attitudes of an audience perceived by the users. In our work, we break down the valence dimension into two sub-dimensions representing the epistemic and affective stances to distinguish the perception of the agreement and of the emotions expressed by virtual characters. (2) On the other hand, we explore the perception of a virtual character's behavior along these different dimensions. Based on the literature, we argue that the perception of the social dimensions is based on behavioral signals specific to each dimension. In other words, we can associate each dimension with a specific behavioral signal (e.g. facial expressions, head movements) that will be used by the user to infer the virtual audience's social attitude. Finally, we formulate these following hypotheses:

H1- Representation of the virtual audience's social attitude.

The valence dimension can be subdivided into an emotional dimension called *affective stance* and an agreement dimension called *epistemic stance*.

H2- Behavioral signals conveying social dimensions. For each dimension, based on the research presented above, we hypothesize that one behavioral signal is predominant in the perception of social attitude:

H2.1- Head Movements. The head movements are the most influential signal for the perception of the virtual audience's *epistemic stance*.

H2.2- Facial Expressions. The facial expressions are the most influential signal for the perception of the virtual audience's *affective stance*.

3 THE EXPERIMENT

In order to investigate the perception of virtual audience behavior, we have conducted a perceptive study in VR in which we asked participants to evaluate the affective and epistemic dimensions of different combinations of nonverbal behavioral signals expressed by a virtual character in a conference room. Our aim was to verify that both the affective and the epistemic dimensions are perceived through the behavior by the users (H1) and that the perception of these dimensions are conveyed by specific behavior signals (H2).

3.1 Virtual Environment and Virtual Character's Nonverbal Behavior

Based on the literature [5, 13], we consider 5 *modalities* (head movements, facial expressions, arms positions, torso positions and gaze directions) and 16 *signals* (e.g. nod, smile) for all these modalities described below. For each of these modalities, we indicate the abbreviation used in Figures 3 and 2:

- Head Movements: Nod (**Nod**), Shake (**Sha**), Tilt (**Til**), No Movement (**NoM**)
- Facial Expressions: Smile (**Smi**), Frown (**Fro**), Doubt (**Dou**), No Expression (**NoE**)
- Arms Position: Fist under the Chin (**Chi**), Arms Crossed (**Cro**), Hands on Thigh (**Thi**)
- Torso Position: Forward (**For**), Backward (**Bac**), Straight (**Str**)
- Gaze Direction: Look at the Speaker (**look**), Look Away from the Speaker (**away**)

From these signals, considering all the possible combinations, we have created a library of 50 animations using the motion capture technology. Each animation corresponds to a combination of several signals, one for each modality¹. We have excluded some combinations not possible physiologically (e.g. the fist under the chin with the torso back) and that are not relevant in our context (e.g. the smile with the head shake). Therefore, an animation corresponds, for example, to the following combination: head tilted, torso back, arms crossed, no facial expression and looking towards the speaker. An animation is made up of *movements* (e.g. nodding, shaking the head and looking away) and *positions* (e.g. tilting, arms crossing, looking towards the speaker etc.). Each animation starts from a neutral position and leads to the target animation in 5 seconds². For example, the virtual character is seated, torso straight, hands on thighs, without head movement or facial expression. It then moves to the final position (e.g. torso back and arms crossed) and performs a head movement if required. To control the viewing time of each animation, we added a user interface (UI) at the end of the animation. This UI was placed on top of the virtual character and allowed participants to choose to replay or to assess the animation. In this study, we first considered a female virtual character aged around 30, and constructed using the Character Creator 4 software (Figure 1). Note that gender and the number of virtual characters

¹Examples of combinations of nonverbal behavioral signals used for some animations: <https://www.youtube.com/playlist?list=PL-2K8UX43M1eeZnJJQxzuzCCXSTaY60iA>

²The duration of the animations corresponds to the duration of the behavior recorded during motion capture.

can have an impact on the perception of the character's social attitude and that the impact of these variables will therefore have to be investigated in more detail in future work [5, 30].



Figure 1: Screenshot of the final position of an animation showing the female virtual character displaying an head tilted, fist under chin, torso leaning forward, no facial expression and a gaze facing the speaker (Til_For_Chi_NoE_look (23)).

3.2 Measures

In order to evaluate the perception of the virtual character's behavior, we asked the users to indicate her/his opinion on different items using a Likert scale of 7 points (from strongly disagree to strongly agree). In a first block, the participants evaluated if the virtual character expressed agreement, disagreement or doubt (e.g. "The virtual character agrees with what is said during the presentation."). In a second block, we asked participants if they perceived the virtual character as happy, unhappy and expressing emotions (e.g. "The virtual character is happy:"). To evaluate the believability of the behavior, we asked the participants if the behavior was realistic and credible (e.g. "The virtual character's behavior is believable.").

3.3 Procedure

In this study, we used a within subjects design. Forty-four unpaid participants, including 20 women, 22 men and 1 non-binary person (34.75 ± 12.27 years) recruited at the University of Aix-Marseille and on social networks, took part in our experiment. We excluded the results of 5 participants due to an interruption in the experiment because of a cable disconnection. We therefore analyzed the results of 39 participants regarding their perception of the 50 animations of the virtual character (Section 3.1).

Firstly, participants were invited to the laboratory and received global information concerning the experiment. We asked participants to sit down and wear the Meta Quest Pro headset. They then fill in a demographic questionnaire, take note of the instructions for the experiment and provide informed consent in VR. After an oral verification of their comprehension, they started the experiment with 3 animations in a familiarization step. These animations were always the same across participants. Next, participants were started

the perceptive study with the 50 animations, with a break at the midway. Each animation lasted 5 seconds. After each animation, the participants had the choice of replaying the animation or moving on to the evaluation of a first block of statements (epistemic stance). If they clicked on continue, we showed them the second block of statements (affective stance and realism). Finally, after the second validation, they accessed a new animation, and we repeated the same process. At the midway, participants had to remove the headset for several minutes, and putted on again to continue the experiment. Note that, we counterbalanced the animation and statements order in each block.

4 RESULTS

In our experiment, we measured the user's perception by considering 8 items, including 3 items for the epistemic stance, 3 for the affective stance and 2 to assess the realism of the animations (Section 3.2). The consistency of the scores has been verified by checking the coherency in the scores of opposite items (agree/disagree, happy/unhappy). Based on Spearman's correlations, we combined the agreement and disagreement items (Spearman's $\rho = -0.60$, p -value < 0.001) corresponding to the epistemic stance and the happy and unhappy items (Spearman's $\rho = -0.40$, p -value < 0.001) corresponding to the affective stance³. The epistemic score is the subtraction of the agreement and disagreement scores, and the affective score the subtraction of the happy and unhappy scores. The two scores range from -6 to 6 and correspond to the dependent variables in our study. Thus, the higher the epistemic score, the more the virtual character is perceived to agree and the higher the affective score, the more the virtual character is perceived as happy. We also computed a believability score ranging from 1 to 7, corresponding to the average of the realistic and credible items (Spearman's $\rho = 0.84$, p -value < 0.001). Note that, on average, the animations were perceived as believable ($M = 5.45 \pm 1.31$).

4.1 Representation of the Virtual Audience's Social Attitude

To answer the first hypothesis (H1), we consider the animation identifier as the independent variable and the epistemic and affective scores as the dependent variables. According to the results of the Shapiro-Wilk test, the distribution of scores differs significantly from normality. We therefore applied the Friedman ANOVA test, a non-parametric test, to study the effect of the independent variable on the dependent variables. The analysis shows an effect of the animations on the epistemic score ($X^2(49) = 1477$; $p < 0.001$) and on the affective score ($X^2(49) = 1410$; $p < 0.001$). Figure 3 and 2 below illustrate the distributions of the scores for each animation. In these figures, we computed three clusters for the two stances, each cluster being illustrated in different colors on the figures. To construct the clusters, we computed the significant differences between the animations using a Nemenyi post-hoc test. Each resulting cluster contains animations that are all significantly different for the animations of the other clusters. For instance, all the animations of the cluster in red in the figures are significantly different

³Due to the lack of correlation between doubt and agreement and between emotion and happiness, these two dimensions were not taken into account when calculating the epistemic and affective stance score.

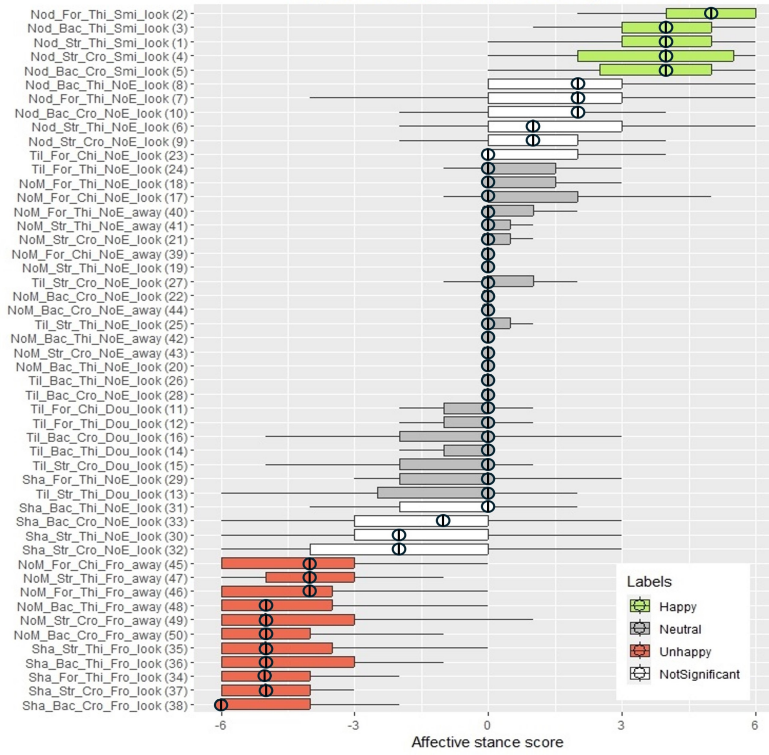


Figure 2: Boxplot of Affective Stance Scores by Animation.

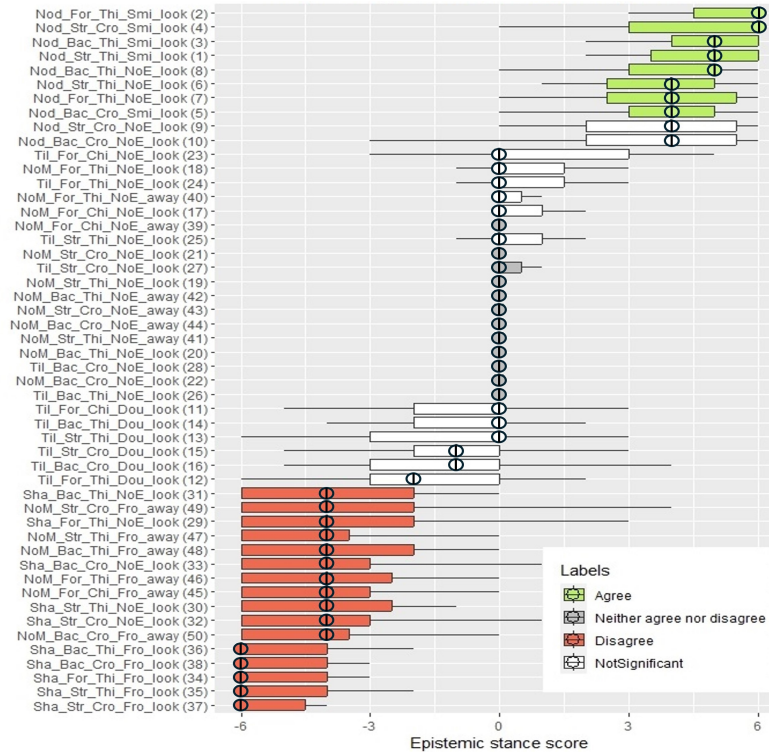


Figure 3: Boxplot of Epistemic Stance Scores by Animation.

| Animations | Epistemic Stance | | | Affective Stance | | |
|---------------------------|------------------|-----|-------|------------------|-----|-------|
| | Md | IQR | Label | Md | IQR | Label |
| Sha_For_Thi_NoE_look (29) | -4 | 2.0 | D | 0 | 2.0 | N |
| Sha_Bac_Thi_NoE_look (31) | -4 | 2.0 | D | 0 | 2.0 | N |
| Sha_Bac_Cro_NoE_look (33) | -4 | 3.0 | D | -1 | 3.0 | N |
| Sha_For_Thi_Fro_look (34) | -6 | 2.0 | D | -5 | 2.0 | U |
| Sha_Str_Thi_Fro_look (35) | -6 | 2.0 | D | -5 | 2.5 | U |
| Sha_Str_Cro_Fro_look (37) | -6 | 1.5 | D | -5 | 2.0 | U |
| Sha_Bac_Cro_Fro_look (38) | -6 | 2.0 | D | -6 | 2.0 | U |

Table 1: Level of affective stance for a given epistemic stance. The seven animations are rated disagree (D) for the epistemic stance, three are rated neutral (N) and four are rated unhappy (U) for the affective stance.

from each animation of the clusters in green and gray. Similarly, all animations in the green cluster are significantly different from each animation in the gray cluster. Thus, for the epistemic character, the green cluster gathers animations in which the virtual character is perceived as agreeing with the speaker ($M = 4.21$, $SD = 1.95$). The gray cluster groups animations in which the virtual character is perceived as neither agreeing nor disagreeing ($M = 0.09$, $SD = 1.43$) and, the red cluster contains animations in which the virtual character is perceived as disagreeing ($M = -4.25$, $SD = 2.10$). For the affective stance, we also identified a green cluster for animations judged to convey happiness ($M = 4.06$, $SD = 1.62$), a gray cluster for animations judged as being neutral ($M = -0.02$, $SD = 1.61$) and a red cluster for animations perceived as unhappy ($M = -4.41$, $SD = 1.83$).

To confirm our first hypothesis, we have to verify that in one of epistemic stance clusters, we have animations belonging to different clusters of affective stance, or, on the contrary, in one of affective stance clusters, we have animations belonging to different clusters of epistemic stances. We first considered the animations of the clusters of the epistemic stance and investigated whether these animations were perceived significantly differently regarding the affective stance. The results show significant differences (Table 1). Indeed, among the animations perceived as conveying disagreement on the epistemic dimension (red cluster, Figure 3) some are perceived as neutral (animations 29, 31 and 33) and others as unhappy (animations 34, 35, 37 and 38) on the affective dimension. No other significant differences were found. Note that we found six animations between the red cluster and the gray cluster that are not significantly different from these clusters, but are different from the green cluster. We checked the doubt score for the six non-significant animations with head tilt, and found an average score of 5.60 out of 7. We can therefore assume that these animations with head tilt and expression of doubt correspond to animations where the virtual character is perceived as doubtful by the participants. Secondly, we consider the animations of the affective stance clusters and investigated whether these animations were perceived significantly differently regarding the epistemic score. Coherently, the results reveal significant differences (Table 2). Finally, according to the combination of nonverbal behavioral signals, participants perceived different epistemic stances for the same affective stance and different affective stances for the same epistemic one. This confirms our first hypothesis (H1).

| Animations | Affective Stance | | | Epistemic Stance | | |
|---------------------------|------------------|-----|-------|------------------|-----|-------|
| | Md | IQR | Label | Md | IQR | Label |
| Sha_For_Thi_NoE_look (29) | 0 | 2.0 | N | -4 | 4.0 | D |
| Til_For_Thi_Dou_look (11) | 0 | 1.0 | N | 0 | 2.0 | N |
| NoM_For_Chi_NoE_look (17) | 0 | 2.0 | N | 0 | 1.0 | N |
| NoM_For_Thi_NoE_look (18) | 0 | 1.5 | N | 0 | 1.5 | N |
| NoM_Str_Thi_NoE_look (19) | 0 | 0.0 | N | 0 | 0.0 | N |
| NoM_Bac_Thi_NoE_look (20) | 0 | 0.0 | N | 0 | 0.0 | N |
| NoM_Str_Cro_NoE_look (21) | 0 | 0.5 | N | 0 | 0.0 | N |
| NoM_Bac_Cro_NoE_look (22) | 0 | 0.0 | N | 0 | 0.0 | N |
| Til_For_Thi_NoE_look (24) | 0 | 1.5 | N | 0 | 1.5 | N |
| Til_Str_Thi_NoE_look (25) | 0 | 0.5 | N | 0 | 1.0 | N |
| Til_Bac_Thi_NoE_look (26) | 0 | 0.0 | N | 0 | 0.0 | N |
| Til_Str_Cro_NoE_look (27) | 0 | 1.0 | N | 0 | 0.5 | N |
| Til_Bac_Cro_NoE_look (28) | 0 | 0.0 | N | 0 | 0.0 | N |
| NoM_For_Chi_NoE_away (39) | 0 | 0.0 | N | 0 | 0.0 | N |
| NoM_For_Thi_NoE_away (40) | 0 | 1.0 | N | 0 | 0.5 | N |
| NoM_Str_Thi_NoE_away (41) | 0 | 0.5 | N | 0 | 0.0 | N |
| NoM_Bac_Thi_NoE_away (42) | 0 | 0.0 | N | 0 | 0.0 | N |
| NoM_Str_Cro_NoE_away (43) | 0 | 0.0 | N | 0 | 0.0 | N |
| NoM_Bac_Cro_NoE_away (44) | 0 | 0.0 | N | 0 | 0.0 | N |

Table 2: Level of epistemic stance for a given affective stance. All animations are rated as neutral (N) for the affective stance, one is rated as disagreeing (D) and eighteen are rated as neither agreeing nor disagreeing (N) for the epistemic stance.

4.2 Behavioral Signals Conveying Social Dimensions

To investigate our second hypothesis (H2), we explore the influence of behavioral signals in the perception of social attitudes. Contrary to the analysis presented in the previous section, in this case, the independent variables are not the animations themselves but the behavioral signals that compose them. Given that all the combinations of behavioral signals are not physiologically possible in terms of experimental design, the plan is incomplete. Note that we excluded the "gaze direction" variable for this analysis because the participants reported that they did not perceive any gaze deviation as we expected. They reported that the character seemed to be looking at the slides behind him. Consequently, to explore the second hypothesis (H2), we used a Bayesian approach. Unlike the frequentist method, this method makes it possible to confirm one hypothesis rather than another thanks to the strength of the evidence [19], measured via the Bayes factor. The Bayes factor (BF_{10}), corresponds to the ratio between the likelihood of the data observed under the alternative hypothesis and the likelihood of the data observed under the null hypothesis [2]. Some authors have tried to quantify the strength of the evidence in order to conclude based on the result of the previous ratio, thanks to different scales [2, 20, 22]. In our study, we chose $BF > 10$ as the threshold based on another study that used data similar to ours [29]. We also used the Natural logarithm $\ln(BF_{10})$ to report the results of BF_{10} when it is greater than 100 [24].

In order to test the two sub-hypotheses (H2-1 and H2-2), we computed two Bayesian models for each stance score, and presented the results concerning our hypotheses in Tables 3 and 4. To test the first sub-hypothesis H2-1, we compared the null model (without

predictors) with the simple model predicting epistemic score using all the modalities (head movements, torso position, arms position and facial expressions). The Bayesian analysis confirms the effects of the simple head movement (model 3.2) and facial expression (model 3.3) models, as shown in the Table 3 with a BF largely superior to the threshold defined above. This approach also allows comparing two models with each other. This means that we can compare the influence of head movements and facial expressions on the perception of epistemic stance. The results show that the ratio of BF between head movements (model 3.2, Table 3) and facial expressions (model 3.3, Table 3) is superior to our threshold ($\ln(BF_{10}) = 232.02$), showing that head movements predominate over facial expressions and over the other signals. *Thus, the head movements are the first and most important signals for assessing the epistemic stance. We also found that facial expressions predominate over the other signals, and therefore represent the second most important signal.* The other simple models did not show sufficient evidence against the null model. We can therefore affirm that the torso position and the arm position have no impact on the assessment of the epistemic score.

To test the second sub-hypothesis (H2-2), using the same method as described above, we compared the null model with the simple model predicting the affective score. The results reported in Table 4 show the effects of the simple facial expressions (model 4.2, Table 4) and head movements (model 4.3, Table 4) models on the perception of the affective stance. By comparing the two models with each other, the results show that facial expressions predominate over the head movements ($\ln(BF_{10}) = 502.86$) and over the other signals. We also found that the head movements predominate over the other signals. *These results show that the facial expression signal is the first and most important signal for assessing affective stance, followed by head movements.* The other simple models also failed to provide sufficient evidence against the null model. We can therefore affirm that the torso position and the arm position have no impact on the assessment of the affective score.

Finally, the head movements and the facial expressions are the most important signals in the perception of the affective and epistemic stances. For the two scores, we then analyzed the effects of the interaction between these two signals on the stances' perception. We also aimed at identifying the best model for explaining the evaluation of each stance. This corresponds to the model with the best BF which is significantly superior to the BF of the other models and which has the lowest number of signals in the model for an equivalent BF .

For this purpose, first, we compared the effect of the models without (model 3.4) and with the interaction (model 3.5) of the two signals (head movements and facial expressions). For the epistemic stance, the BF ratio between these two models, $\ln(BF_{10}) = 75.00$, is in favor of the interaction model (model 3.5). Based on the scale defined by Jeffreys [20], the results provide strong evidence that the model with the interaction (model 3.5) is better than the one without the interaction (model 3.4). Therefore, we can confirm that there is an interaction between head movements and facial expressions in the perception of the epistemic stance. Among the possible models, the best model is the model combining head movements, arms position, facial expressions and the interaction of head movements and facial expressions (model 3.5 + arms position) with $BF_{10} = 1.63 \times 10^{552}$. However, this model is not better than the same model

| Models | P(M) | P(M data) | $\log(BF_M)$ | $\log(BF_{10})$ | BF_{10} | error % |
|-----------------------------|------|-------------------------|--------------|-----------------|------------------------|---------|
| (3.1) Null model (incl. ID) | 0.01 | 0.00 | -1266.98 | 0.00 | 1.00 | |
| (3.2) H. | 0.01 | 1.16×10^{-167} | -379.27 | 887.71 | 3.39×10^{385} | 0.76 |
| (3.3) Exp. | 0.01 | 2.01×10^{-268} | -611.28 | 655.70 | 5.81×10^{284} | 0.72 |
| (3.4) H. + Exp. | 0.01 | 2.91×10^{-34} | -72.11 | 1194.87 | 8.56×10^{518} | 0.92 |
| (3.5) H. + Exp. + H. * Exp. | 0.01 | 1.10 | 2.96 | 1269.83 | 3.20×10^{551} | 2.19 |

Table 3: Epistemic stance prediction using Bayes factor ANOVA: The table reports the results for the simple models of head movements (H.), facial expressions (Exp.), their association (+) and their interaction (*).

| Models | P(M) | P(M data) | $\log(BF_M)$ | $\log(BF_{10})$ | BF_{10} | error % |
|-----------------------------|------|-------------------------|--------------|-----------------|------------------------|---------|
| (4.1) Null model (incl. ID) | 0.01 | 0.00 | -1266.98 | 0.00 | 1.00 | |
| (4.2) Exp. | 0.01 | 5.77×10^{-76} | -168.13 | 1027.82 | 2.41×10^{446} | 0.59 |
| (4.3) H. | 0.01 | 2.41×10^{-294} | -670.97 | 524.99 | 9.80×10^{227} | 1.13 |
| (4.4) H. + Exp. | 0.01 | 4.13×10^{-14} | -25.71 | 1170.25 | 1.80×10^{508} | 1.10 |
| (4.5) H. + Exp. + H. * Exp. | 0.01 | 2.62×10^{-4} | -3.13 | 1192.82 | 1.24×10^{518} | 7.93 |

Table 4: Affective stance prediction using Bayes factor ANOVA: The table reports the results for the simple models of head movements (H.), facial expressions (Exp.), their association (+) and their interaction (*).

without the arm positions (model 3.5) with $BF_{10} = 3.20 \times 10^{551}$ because when we compare these two models, we obtain $BF_{10} = 5.11$ in favor of the more complex model which is inferior to the fixed threshold. We can confirm that the model 3.5 is the best model to assess the epistemic stance score.

Concerning the affective stance, we applied the same methods. We compared the effect of the model without interaction (model 4.4, Table 4) and with interaction (model 4.5, Table 4) on the affective score. In this case, we found a Bayes factor of $\ln(BF_{10}) = 22.66$ in favor of the interaction model (model 4.5). We can confirm the presence of an interaction between the facial expressions and the head movements to assess the affective stance. Among the possible models, the best model to assess the affective score is the model combining head movement, torso position, facial expressions and the interaction between head movements and facial expressions (model 4.5 + torso position, Table 4). However, this model is not significantly different from the model with arms position (model 4.5 + arm position) which is the second-best model with a Bayes factor of $BF_{10} = 8.38 \times 10^{520}$ and the best model with the fewest signals.

Finally, the results of our experiment confirm our first hypothesis (H1): based on the combination of nonverbal behavioral signals, participants perceived different epistemic stances for the same affective stance and different affective stances for the same epistemic stance. Moreover, each stance is assessed mainly based on specific nonverbal signal, this confirms our second hypothesis (H2). The head movements are mainly involved in assessing the epistemic dimension, and the facial expressions in assessing the affective dimension.

5 DISCUSSION

In this work, we question the valence-arousal model used in the literature [5, 13, 16, 21] to simulate the social attitudes of a virtual audience. The valence dimension covers two different virtual audience's stance dimensions that should be considered separately.

Consequently, we propose to subdivide the valence dimension into 2 sub-dimensions characterizing, on the one hand, the epistemic stance (i.e. the virtual character's agreement with the speaker's presentation), and, on the other hand, the affective stance (i.e. the emotions expressed in reaction to the speaker's presentation). To demonstrate the relevancy of this bi-dimensional representation of the valence, in this article, we have presented a user perceptual study of virtual character animations considering these two dimensions.

Epistemic Stance. The results of the study confirm our first hypothesis (H1). Indeed, the virtual character is perceived as agreeing, neither agreeing nor disagreeing or disagreeing depending on the combination of displayed nonverbal behavioral signals, and for the animations judged to convey disagreement, participants perceived different affective stances. Among the animations in which the character was perceived as disagreeing, participants perceived a neutral emotion when the character had no facial expression and a negative emotion (unhappiness) when the character frowned and had a sad mouth. Therefore, the nonverbal behaviors may convey a specific epistemic stance but different affective stances. This provides some evidence in favor of a subdivision of the valence dimension into two sub-dimensions: the epistemic stance and the affective stance dimension.

The results confirm, moreover, our second hypothesis (H2). Indeed, to assess the epistemic stance, participants mainly used the head movements (H2.1). The head shake signals in the animations were all rated negatively for agreement, indicating, as in the literature, that the head shake is a cue showing the negative opinion of a virtual character [5, 13, 16]. The head nodding in the animations were positively evaluated for agreement, which indicates, as in the literature, that head nodding is a cue showing the positive opinion of a virtual character [5, 13, 16]. Regarding head tilt, we did not find any animation with a head tilt that was significantly different from the others. However, it appears that these animations were slightly negatively rated, but less so than the animations rating as disagreement. This is consistent with the literature indicating that head tilt is a negative epistemic stance showing speaker uncertainty [14] and is associated with a negative valence [13]. We also note that the animations rated neither agree nor disagree correspond mainly to animations with a virtual character without head movements.

Affective Stance. The results concerning the affective stance confirm the first hypothesis (H1). The virtual character has been perceived with different affective stances depending on its nonverbal behavior. Some animations represent a happy virtual character, others a neutral or unhappy virtual character. In addition, for animations judged to be neutral, we also observed different epistemic stances. Among the animations judged as neutral on the affective dimension, we found that eighteen animations showed a character who neither agreed nor disagreed, with no head movement or with a head tilt, and one animation showed a character who disagreed with a head shake. This finding contributes to confirm our first hypothesis. For the same affective stance, a virtual character may convey different epistemic stances.

The results moreover confirm the second sub-hypothesis (H2.2). Indeed, the participants mainly used the facial expressions to assess the affective stance. The frown signals in the animations are all rated negatively for unhappiness, indicating, as in the literature, that the

frown is a cue showing the negative emotion of a virtual character [5, 13, 16]. The smile signals in animations is all rated positively for happiness, indicating, as in the literature, that the smile is a strong cue of a positive emotion [5, 13, 16]. All the animations with an expression of doubt belong to the gray cluster and are judged to be neutral compared to the other animations, judged to be happy and unhappy. This is consistent with the results of Etienne et al. [13] showing that raised eyebrows, which are close to a doubt expression, reflect doubt. Moreover, doubt can be viewed as a kind of neutral emotion. The rest of the animations judged as neutral correspond to animations in which the virtual character has no specific facial expression. If we were unable to associate the head tilt with the label neither agree nor disagree, we examined whether participants perceived this as doubt, and it was. This might suggest that doubt is not really a neutral state between 'agree' and 'disagree'. A head tilt could have a negative connotation, but not as much as a head shake. However, the facial expression of doubt is closer to neutral emotion than to frowning and smiling.

Among the results of this study, we also showed the absence of influence of arm and torso position on the perception of epistemic and affective stances. Indeed, the most important nonverbal behavioral signals for judging the agreement and emotions expressed by the virtual character are head movements and facial expressions.

6 CONCLUSION

In this article, through a user perceptual study, we aimed to show that the valence dimension of the valence-arousal model should be divided into two sub-dimensions to describe the nonverbal behavior of the characters conveying social attitudes to a virtual audience. The results of our experiment show that the valence dimension can be subdivided into an emotional dimension reflecting the affective stance and an agreement dimension corresponding to the epistemic stance. Indeed, a virtual character can convey a combination of different affective and epistemic stance through its nonverbal behavior. Moreover, a particular behavioral signal is associated to each dimension. To assess the epistemic stance, the main signal used by the users is the head movements, while to assess the affective stance, the main signal is the facial expressions. We also found that doubt cannot be considered as a neutral state between agreement and disagreement because it seems to be assessed as slightly negative.

The subdivision of the valence into the two proposed dimensions could allow creating more than the six social attitude identify in the literature if we consider the combination of the level of engagement (engage or disengage), the epistemic stance (agree, neither agree nor disagree and disagree) and the affective stance (happy, neutral and unhappy). In a next step, we plan to study the social attitudes resulting from this three-dimensional model.

This study presents certain limitations. We studied the perception of the nonverbal behaviors of a single female virtual character. Appearance, and more specifically the gender of the virtual character, may influence the user's perception [30], indicating the need to confirm these results by studying this variable. Furthermore, given that the perception of an audience is not the average perception of each virtual character [5], a perceptual study with an entire audience should be conducted. Last but not least, in our study,

participants were not involved in a public speaking task. However, the user's engagement as a speaker could also influence his or her perception of the virtual audience's epistemic and affective stances.

ACKNOWLEDGMENTS

This research was funded under the ANR REVITALISE grant ANR-21-CE33-0016-02.

REFERENCES

- [1] Lígia Batrinca, Giota Stratou, Ari Shapiro, Louis-Philippe Morency, and Stefan Scherer. 2013. Cicero-towards a multimodal virtual audience platform for public speaking training. In *Intelligent Virtual Agents: 13th International Conference, IVA 2013, Edinburgh, UK, August 29-31, 2013. Proceedings 13*. Springer, 116–128.
- [2] Daniel J Benjamin and James O Berger. 2019. Three recommendations for improving the use of p-values. *The American Statistician* 73, sup1 (2019), 186–191.
- [3] Pio E Ricci Bitti, Luisa Bonfiglioli, Paolo Melani, Roberto Caterina, and Pierluigi Garotti. 2014. Expression and communication of doubt/uncertainty through facial expression. *Ricerche di Pedagogia e Didattica. Journal of Theories and Research in Education* 9, 1 (2014), 159–177.
- [4] Massimo Chindamo, Jens Allwood, and Elisabeth Ahlsén. 2012. Some suggestions for the study of stance in communication. In *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing*. IEEE, 617–622.
- [5] Mathieu Chollet and Stefan Scherer. 2017. Perception of virtual audiences. *IEEE computer graphics and applications* 37, 4 (2017), 50–59.
- [6] Mathieu Chollet, Giota Stratou, Ari Shapiro, Louis-Philippe Morency, and Stefan Scherer. 2014. An interactive virtual audience platform for public speaking training. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*. 1657–1658.
- [7] Bianca Malaika Ciuffani. 2017. *Non-verbal Communication and Leadership: the impact of hand gestures used by leaders on follower job satisfaction*. B.S. thesis. University of Twente.
- [8] Fiona Dermody and Alistair Sutherland. 2018. Multimodal systems for public speaking: a case in support of a positive computing approach. In *Proceedings of the 2nd International Conference on Computer-Human Interaction Research and Applications-Volume 1: CHIRA, Vol. 1*. Scitepress, 170–175.
- [9] John P Doody and Peter Bull. 2011. Asperger's syndrome and the decoding of boredom, interest, and disagreement from body posture. *Journal of nonverbal behavior* 35 (2011), 87–100.
- [10] Mariia Dubiago, Sandra Poeschl, and Nicola Doering. 2017. Influence of Simulation Fidelity on Perceived Simulation Realism—An Exploratory Study on a Virtual Public Speaking Training Application. *Annu. Rev. CyberTher. Telemed* 15 (2017), 154–159.
- [11] Meriem El-Yamri, Alejandro Romero-Hernandez, Manuel Gonzalez-Riojo, and Borja Manero. 2019. Designing a VR game for public speaking based on speakers features: a case study. *Smart Learning Environments* 6, 1 (2019), 12.
- [12] Paul MG Emmelkamp, Katharina Meyerbröker, and Nexhmedin Morina. 2020. Virtual reality therapy in social anxiety disorder. *Current psychiatry reports* 22 (2020), 1–9.
- [13] Elodie Etienne, Anne-Lise Leclercq, Angélique Remacle, Laurence Dessart, and Michaël Schyns. 2023. Perception of avatars nonverbal behaviors in virtual reality. *Psychology & Marketing* 40, 11 (2023), 2464–2481.
- [14] Farina Freigang, Sören Klett, and Stefan Kopp. 2017. Pragmatic multimodality: Effects of nonverbal cues of focus and certainty in a virtual human. In *Intelligent Virtual Agents: 17th International Conference, IVA 2017, Stockholm, Sweden, August 27-30, 2017, Proceedings 17*. Springer, 142–155.
- [15] Yann Glémarec, Jean-Luc Lugin, Anne-Gwenn Bosser, Cédric Buche, and Marc Erich Latoschik. 2022. Controlling the stage: a high-level control system for virtual audiences in Virtual Reality. *Frontiers in Virtual Reality* 3 (2022), 876433.
- [16] Yann Glémarec, Jean-Luc Lugin, Anne-Gwenn Bosser, Aryana Collins Jackson, Cédric Buche, and Marc Erich Latoschik. 2021. Indifferent or enthusiastic? virtual audiences animation and perception in virtual reality. *Frontiers in Virtual Reality* 2 (2021), 666232.
- [17] Bin Han and Gerard Jounghyun Kim. 2021. Audiencemr: Extending the local space for large-scale audience with mixed reality for enhanced remote lecturer experience. *Applied Sciences* 11, 19 (2021), 9022.
- [18] Michał Jakubowski, Marcin Wardaszko, Anna Winniczuk, Błażej Podgórski, and Małgorzata Cwil. 2019. Cicero VR-Public Speaking Training Tool and an Attempt to Create Positive Social VR Experience. In *Virtual, Augmented and Mixed Reality. Applications and Case Studies: 11th International Conference, VAMR 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26–31, 2019, Proceedings, Part II 21*. Springer, 297–311.
- [19] Andrew F Jarosz and Jennifer Wiley. 2014. What are the odds? A practical guide to computing and reporting Bayes factors. *The Journal of Problem Solving* 7, 1 (2014), 2.
- [20] Harold Jeffreys. 1998. *The theory of probability*. OUP Oxford.
- [21] Ni Kang, Willem-Paul Brinkman, M Birna Van Riemsdijk, and Mark Neerincx. 2016. The design of virtual audiences: noticeable and recognizable behavioral styles. *Computers in Human Behavior* 55 (2016), 680–694.
- [22] Robert E Kass and Adrian E Raftery. 1995. Bayes factors. *Journal of the american statistical association* 90, 430 (1995), 773–795.
- [23] Emmanuelle P Kleinlogel, Marion Curdy, João Rodrigues, Carmen Sandi, and Marianne Schmid Mast. 2021. Doppelgänger-based training: Imitating our virtual self to accelerate interpersonal skills learning. *PLoS one* 16, 2 (2021), e0245960.
- [24] Nicolas Mascaret, Olivier Vors, Tanguy Marqueste, Rémy Casanova, and François Cury. 2019. Social support from evaluative familiar persons—a buffer against stress? Preliminary evidence of neuroendocrine and psychological responses. *Anxiety, Stress, & Coping* 32, 5 (2019), 534–544.
- [25] Emma Otta, Beatriz Barcellos Pereira Lira, Nadia Maria Delevati, Otávio Pimentel Cesar, and Carla Salati Guirello Pires. 1994. The effect of smiling and of head tilting on person perception. *The Journal of psychology* 128, 3 (1994), 323–331.
- [26] Fabrizio Palmas, Jakub Cichor, David A Plecher, and Gudrun Klinker. 2019. Acceptance and effectiveness of a virtual reality public speaking training. In *2019 IEEE international symposium on mixed and augmented reality (ISMAR)*. IEEE, 363–371.
- [27] Fabrizio Palmas, Ramona Reinelt, Jakub E Cichor, David A Plecher, and Gudrun Klinker. 2021. Virtual reality public speaking training: Experimental evaluation of direct feedback technology acceptance. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE, 463–472.
- [28] David-Paul Pertaub, Mel Slater, and Chris Barker. 2002. An experiment on public speaking anxiety in response to three different types of virtual audience. *Presence* 11, 1 (2002), 68–78.
- [29] Jakub Procházka, Daniel Kerekeš, et al. 2017. Influence of autonomy and task significance on engagement: Experiment. In *Trends and Issues in Interdisciplinary Behavior and Social Science: Proceedings of the 5th International Congress on Interdisciplinary Behavior and Social Science (ICIBSoS 2016), 5-6 November 2016, Jogjakarta, Indonesia*. CRC Press, 21–24.
- [30] Marion Ristorcelli, Emma Gallego, Kevin Nguy, Jean-Marie Pergandi, Rémy Casanova, and Magalie Ochs. 2023. Investigating the impact of a virtual audience's gender and attitudes on a human speaker. In *Companion Publication of the 25th International Conference on Multimodal Interaction*. 363–367.
- [31] Christian Seufert, Sebastian Oberdörfer, Alice Roth, Silke Grafe, Jean-Luc Lugin, and Marc Erich Latoschik. 2022. Classroom management competency enhancement for student teachers using a fully immersive virtual classroom. *Computers & Education* 179 (2022), 104410.
- [32] Carolin Straßmann, Astrid Rosenthal von der Pütten, Ramin Yaghoobzadeh, Raffael Kaminski, and Nicole Krämer. 2016. The effect of an intelligent virtual agent's nonverbal behavior with regard to dominance and cooperativity. In *Intelligent Virtual Agents: 16th International Conference, IVA 2016, Los Angeles, CA, USA, September 20–23, 2016, Proceedings 16*. Springer, 15–28.
- [33] Robin E Sülter, Paul E Ketelaar, and Wolf-Gero Lange. 2022. SpeakApp-Kids! Virtual reality training to reduce fear of public speaking in children—A proof of concept. *Computers & Education* 178 (2022), 104384.
- [34] BK Wiederhold and G Riva. 2012. Designing virtual audiences for fear of public speaking training—an observation study on realistic nonverbal behavior. *Annu. Rev. Cybertherapy Telemed. Adv. Technol. Behav. Soc. Neurosci* 181 (2012), 218.
- [35] Haipeng Zeng, Xingbo Wang, Yong Wang, Aoyu Wu, Ting-Chuen Pong, and Huamin Qu. 2022. GesturElen: Visual analysis of gestures in presentation videos. *IEEE Transactions on Visualization and Computer Graphics* (2022).