



**HAL**  
open science

## Indifferent or Enthusiastic? Virtual Audiences Animation and Perception in Virtual Reality

Yann Glémarec, Jean-Luc Lugin, Anne-Gwenn Bosser, Aryana Zaleh Collins  
Jackson, Cédric Buche, Marc Erich Latoschik

► **To cite this version:**

Yann Glémarec, Jean-Luc Lugin, Anne-Gwenn Bosser, Aryana Zaleh Collins Jackson, Cédric Buche, et al.. Indifferent or Enthusiastic? Virtual Audiences Animation and Perception in Virtual Reality. *Frontiers in Virtual Reality*, 2021, 2, pp.666232. 10.3389/frvir.2021.666232 . hal-03244205

**HAL Id: hal-03244205**

**<https://hal.science/hal-03244205>**

Submitted on 2 Jul 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.



Distributed under a Creative Commons Attribution 4.0 International License



# Indifferent or Enthusiastic? Virtual Audiences Animation and Perception in Virtual Reality

Yann Glémarec<sup>1,2\*</sup>, Jean-Luc Lugrin<sup>2</sup>, Anne-Gwenn Bosser<sup>1</sup>, Aryana Collins Jackson<sup>1</sup>, Cédric Buche<sup>3</sup> and Marc Erich Latoschik<sup>2</sup>

<sup>1</sup>Lab-STICC, ENIB, CNRS UMR 6285, Brest, France, <sup>2</sup>Human-Computer Interaction Group, University of Würzburg, Würzburg, Germany, <sup>3</sup>CROSSING, ENIB, CNRS IRL 2010, Adelaide, SA, Australia

In this paper, we present a virtual audience simulation system for Virtual Reality (VR). The system implements an audience perception model controlling the nonverbal behaviors of virtual spectators, such as facial expressions or postures. Groups of virtual spectators are animated by a set of nonverbal behavior rules representing a particular audience *attitude* (e.g., indifferent or enthusiastic). Each rule specifies a nonverbal behavior category: *posture*, *head movement*, *facial expression* and *gaze direction* as well as three parameters: *type*, *frequency* and *proportion*. In a first user-study, we asked participants to pretend to be a speaker in VR and then create sets of nonverbal behaviour parameters to simulate different attitudes. Participants manipulated the nonverbal behaviours of single virtual spectator to match a specific levels of engagement and opinion toward them. In a second user-study, we used these parameters to design different types of virtual audiences with our nonverbal behavior rules and evaluated their perceptions. Our results demonstrate our system's ability to create virtual audiences with three types of different perceived attitudes: *indifferent*, *critical*, *enthusiastic*. The analysis of the results also lead to a set of recommendations and guidelines regarding attitudes and expressions for future design of audiences for VR therapy and training applications.

**Keywords:** virtual reality, perception, nonverbal behavior, interaction, virtual agent, virtual audience

## OPEN ACCESS

### Edited by:

Katja Zibrek,  
Inria Rennes—Bretagne Atlantique  
Research Centre, France

### Reviewed by:

Rebecca Fribourg,  
Trinity College Dublin, Ireland  
Brian Ravenet,  
Université Paris-Saclay, France

### \*Correspondence:

Yann Glémarec  
yann.glemarec@uni-wuerzburg.de

### Specialty section:

This article was submitted to  
Technologies for VR,  
a section of the journal  
Frontiers in Virtual Reality

**Received:** 09 February 2021

**Accepted:** 18 May 2021

**Published:** 01 June 2021

### Citation:

Glémarec Y, Lugrin J-L, Bosser A-G,  
Collins Jackson A, Buche C and  
Latoschik ME (2021) Indifferent or  
Enthusiastic? Virtual Audiences  
Animation and Perception in  
Virtual Reality.  
Front. Virtual Real. 2:666232.  
doi: 10.3389/frvir.2021.666232

## 1 INTRODUCTION

Virtual Reality (VR) systems have increasingly been used for social simulation applications, such as those for training or therapy. Such systems often require the VR environment to be populated with groups of virtual spectators, whether for public speaking skills (Batrinca et al., 2013; Chollet et al., 2014) for audience management training (Hayes et al., 2013; Lugrin et al., 2016; Fukuda et al., 2017; Shernoff et al., 2020), or forms of social anxiety disorders (Wallach et al., 2009; Anderson et al., 2013; Kahlon et al., 2019). The virtual spectators (i.e., virtual agents who watch an activity without taking part) populating these environments are called a Virtual Audience (VA) and the models underlying their simulated behaviors are of primary importance for users training and therapy outcomes.

Virtual Reality promises environments which can be dynamically controlled at a level that would be mostly unfeasible or unsafe during *in vivo* simulation, in a real environment such as a lecture room. The need for such fine control is made especially evident in applications such as exposure therapy, which consists of repeatedly exposing a patient to varying degrees of a feared stimuli in order

to modify a behavioral or cognitive response (Rothbaum et al., 2000; Anderson et al., 2005). For this promise to hold for VR applications where the social aspect is key such as public speaking anxiety treatment, a fine control of the audience attitude is paramount for rooting the user in the virtual scene and providing training and therapeutic adaptive environments.

The effectiveness of VR applications for both therapy and training strongly relies on the phenomenon of Presence which is known as the feeling of “being here” or to be the moment when “there is successful substitution of real sensory data by virtually generated sensory data” (Slater et al., 2009). Virtual training systems research suggests that Presence achieved through an interactive virtual environment significantly improves learning effectiveness (Messinis et al., 2010). Whilst Presence is achievable without VR, several studies show that VR training systems produce a stronger sense of presence than systems using desktop screens (Shu et al., 2019; Delamarre et al., 2020).

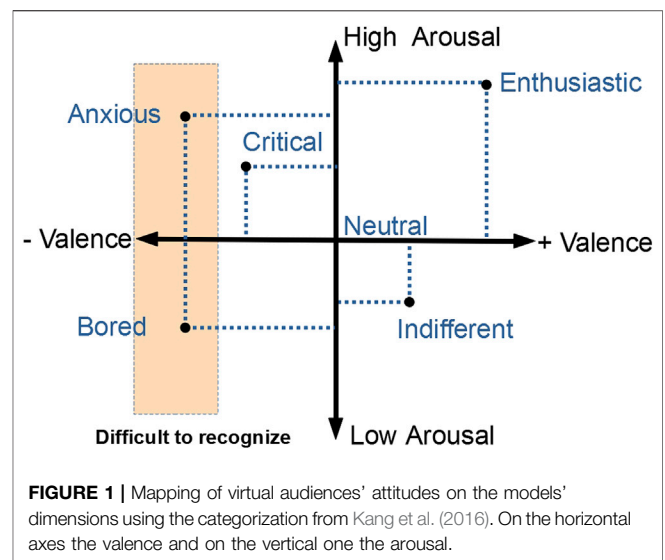
In this paper, we describe a rule-based system enabling the control of various VA behavior simulation. The system’s underlying model is based on recommendations from previous research on audience user-perception for generated desktop videos (Kang et al., 2016; Chollet and Scherer, 2017). We also describe and discuss the results of two user evaluations for our system. We first evaluate individual virtual agents’ nonverbal behaviors according to two dimensions, the valence and the arousal, against the recommendations of existing behavior models. Results suggest that postures’ openness is not significantly associated with the valence. Then, the overall VA’s attitude perception is evaluated depending on the same dimensions. Results reveal users’ difficulties to make a distinction between levels of arousal when the valence is negative. All of these results help us to provide guidelines and recommendations for the design of VAs for virtual reality.

## 2 RELATED WORKS

### 2.1 Virtual Audience Behavior Models

Existing VAs have relied on a wide variety of approaches. The use of cognitive models such as Pleasure-Arousal-Dominance (Heudin, 2007), Appraisal (Marsella and Gratch, 2002), or Valence-Arousal (Chollet et al., 2014) are used in many cases and have led to different implementations. A straightforward representation of these models is that continuous models (Valence-Arousal, Pleasure-Arousal-Dominance) map an individual’s emotional states along dimensions (Mehrabian, 1996) whereas discrete models describe fixed emotion like the basic emotions from (Ekman, 1999). In appraisal theory, models state the importance of the evaluation and the interpretation of an event to explain an individual’s emotions (Roseman, 1991).

All of these varying cognitive models are used for the design of believable nonverbal behaviors, such as facial-expressions, postures or gestures (Heudin, 2004; Gebhard et al., 2018). However, behavior models taking into account group behaviors or even audience behaviors are fewer in number. Models such as these have to include the overall perception of these behaviors and potential interactions between them, which



makes these models more complex to evaluate. As a result, studies use different approaches to investigate nonverbal behaviors perception. One solution is to use crowd-sourcing to get large samples of users who design the behaviors themselves (Chollet and Scherer, 2017). Other models are built by analyzing video records to get a representative corpus and identifying patterns with a statistical approach (Kang et al., 2016), or with user evaluations and past results from the literature (Pelachaud, 2009; Fukuda et al., 2017; Hosseinpanah et al., 2018). Another common method is using domain experts’ and users’ knowledge when related to a specific context (Lugrin et al., 2016; Kahlon et al., 2019). However, there is a severe deficiency in audience behavior research which means that building a VA behavior model upon existing results becomes difficult (Wiederhold and Riva, 2012).

With regard to perception of VA behavior, Kang et al. (2016) provide two dimensions to describe how audiences are perceived. The dimensions, the valence and the arousal, were significantly recognized in VR. They used records of real audiences displaying specific attitudes like *bored* or *interested* to encode virtual spectators’ behaviors. They applied these behaviors to VAs and then investigated how users perceived them. The valence-arousal pair can be used to match with different VA’s attitudes (Kang et al., 2016). This mapping can be useful for both describing perceived attitudes and designing VAs, the **Figure 1** reports these attitudes on the valence and arousal axis. In line with this research, Chollet and Scherer (2017) propose a model based on the same dimensions (Valence-Arousal). Their studies offer a detailed model on how nonverbal behaviors are individually perceived and associated with one of these dimensions according to users perception. Moreover, this model provides rich details on how the VA will be perceived according to its design (agents’ location or the number of agents displaying a targeted behavior). Our current VA’s behaviors are grounded in this model which allows us to provide a variety of audience attitudes along the valence-arousal pair. However, their results and recommendations are based on a crowd sourced study

relying on VAs videos displayed on desktop screens and not directly in VR.

## 2.2 Perception in Virtual Reality

If training and therapeutic VR applications stand in need of VAs to enhance training outcomes and treatment results, their behavior needs to be accurate in order to correctly display the targeted attitude (the feared stimuli or the training scenario). The aforementioned audience model for desktop from Chollet and Scherer (2017) provides a precise description of the perceived nonverbal audience behaviors according to the valence and arousal dimensions. With this detailed model we are able to establish nonverbal behavior rules for the VAs. These two dimensions are depicted as the opinion toward the speech or the user's avatar (Valence) and the engagement towards the speech or the user's avatar (Arousal). Their recommendations provide a list of nonverbal behaviors such as gaze direction, gaze away frequency, postures in terms of openness and proximity, type of head movements', and type and frequency of facial expressions. On top of these behaviors the model gives more insight into the audience design and its perception. Chollet and Scherer (2017) investigate how many virtual spectators need to be manipulated to allow the user to perceive the targeted attitude. According to Chollet and Scherer (2017)'s work, the gaze away frequency is the most influential behavior for the perception of the audience's engagement. The head movement is the most influential behavior for the perception of the audience's opinion. This model was evaluated for an audience of 10 agents placed on two rows. For this configuration, the location of individual agents had no significant effect on the users' identification rate. Finally, the perceived audience's attitude varied depending on the number of individual agents displaying a specific behavior. For a given VA of 10 agents, three agents showing a negative behavior will likely trigger the perception of a negative audience, whereas the threshold goes up to six agents displaying a positive behavior for the perception of a positive audience. Audience engagement may be perceived from at least four virtual spectators displaying engagement among 10 agents. Nonetheless, three different facial expressions, two types of head movements and six different postures is sufficient to allow a user to perceive different audience attitudes. These behaviors remain generic and this model was not evaluated for a specific context.

However, studies seem to indicate various differences between VR head mounted displays and traditional computer screens, as was the case in (Chollet and Scherer, 2017). Technical differences may alter users' perception. As a matter of fact, when using an *egocentric* point of view (VR), the position, the orientation, and movements of objects are defined according to the body whilst with an *exocentric* point of view (desktop screens), it is not affected by the body location (Bowman et al., 2004). Hence, a VR environment where the user's position is dynamic may be different from a static exocentric one. Furthermore, the fact that the user's body is included in the environment may alter the VA perception compared to a third person viewpoint. VR embodiment provides a sensation of immersion, given that movement tracking, latencies, field of view, audio and haptic feedback are issued. Different studies state that a low latency with

head tracking and a wide stereoscopic field of view enhance the feeling of presence and the performances for different tasks in VR (Arthur et al., 1993; Hale and Stanney, 2006; Lee et al., 2010; Lugrin et al., 2013) as well as a good body tracking (Cummings and Bailenson, 2016). Moreover, these technical prerequisites, if added to a better feeling of immersion and realistic interactions with the virtual environment or other users can significantly enhance not only performances related to VR tasks but also communications between users (Narayan et al., 2005). Recent studies on social VR which exploits rich social signals and behavior patterns explored how to leverage these VR requirements for the feeling of co-presence as well as the interactions and the immersion by adding co-located agents and an embodied avatar for the user to interact with the virtual environment (Latoschik et al., 2019). This shift in technology raises the question on the potential difference in audience behavior perception between VR devices and traditional desktop screens. In terms of social-presence, defined as the "sense of being with another" (Biocca et al., 2003), VR environments populated with virtual spectators produce a greater feeling of co-presence compared to desktop environments (Guimarães et al., 2020).

In order to highlight dissimilarities in users' perception between VR and desktop contexts, as well as to provide guidelines to the VR community for VA design, we chose to adhere to the protocol used by Chollet and Scherer (2017): during the first phase, users were asked to design nonverbal behaviors according to given values of valence and arousal. During the second phase of the experiment, a different set of users were asked to rate the audiences' behaviors in terms of valence and arousal. Difference in methodology has arisen from the difference of context: Chollet and Scherer (2017) used a crowd-sourced study, which was difficult to achieve in VR, so we have run a user study instead. This allows us to compare nonverbal behaviors' perception and audiences' perceived attitudes to those obtained in the original crowd-sourced evaluation. Comparison can highlight dissimilarities in users' perception and allow use to provide guideline to the VR community for VA design.

## 3 VIRTUAL AUDIENCE GENERATION SYSTEM

Our system allows for the simulation and control of VAs by managing the virtual spectators' behavior, as the combination of the agents' attitude influences the audience perception. The simulation is built upon nonverbal behavior rules for both single agents and entire audiences. Our implementation relies on the Unreal®Engine 4, which is a high-end development framework widely used for 3D games and VR software. To facilitate its use and future distribution, the entire system is implemented as an open sourced plugin<sup>1</sup> that we designed and

<sup>1</sup><http://hci.uni-wuerzburg.de/projects/virtual-audiences/>

**TABLE 1 |** Example of a rules set for an Enthusiastic and critical attitudes. The frequency parameter used are the same as in our user study.

Enthusiastic				
Arousal		Very high		
Valence		Very positive		
Rules set	Type	Frequency	Proportion	
<i>Rule<sub>Posture</sub></i>	Lean forward chin on fist	Always (1.0)	20%	
<i>Rule<sub>Posture</sub></i>	Upright hands on laps	Always (1.0)	20%	
<i>Rule<sub>Posture</sub></i>	Forward hands together	Always (1.0)	20%	
<i>Rule<sub>FacialExpression</sub></i>	Smile	Most of the time (0.75)	60%	
<i>Rule<sub>HeadMovement</sub></i>	Nod	Most of the time (0.75)	50%	
<i>Rule<sub>gaze</sub></i>	Sideways	Rarely (0.25)	10%	
Critical				
Arousal		Medium/High		
Valence		Negative		
<i>Rule<sub>Posture</sub></i>	Backward arms crossed	Always (1.0)	20%	
<i>Rule<sub>Posture</sub></i>	Upright self hold	Always (1.0)	20%	
<i>Rule<sub>FacialExpression</sub></i>	Frown	Often (0.75)	60%	
<i>Rule<sub>HeadMovement</sub></i>	Shake	Sometimes (0.5)	40%	
<i>Rule<sub>gaze</sub></i>	Downward	Sometimes (0.5)	20%	
<i>Rule<sub>gaze</sub></i>	Sideways	Sometimes (0.5)	20%	

implemented which is compatible with most recent versions of the Unreal®Engine.

### 3.1 Virtual Reality Adaption and Nonverbal Behavior Rules

We formalised the recommendations from Chollet and Scherer (2017) into our VR implementation in the form of a set of user-customizable rules. A rule is a series of parameters describing a nonverbal behavior that can be applied to the audience’s virtual agents. An audience’s attitude is therefore specified through rules. Rules are divided into categories corresponding to the model’s nonverbal behaviors (posture, gaze, head movement and facial expression). A rule follows the following format:

$$rule_x (Type, Frequency, Proportion) \tag{1}$$

where *x* is the nonverbal behavior category of the rule (e.g., posture, gaze), *Type* a pre-defined parameter characterising the nonverbal behavior in the category, *Frequency* how often the behavior is displayed for each active agent, and *Proportion* the number of agents in the audience which will be actively displaying the behaviour. An example of a rule for the facial expression category would be:

$$rule_{FacialExpression} (Smile, 0.5, 0.7) \tag{2}$$

This can be read as 70% of the agents smile 50% of a given period. The system combines any set of given rules into audience animations. Hence, the overall VA’s attitude is an aggregate of each nonverbal behaviors which are allocated to the virtual agents. The rest of the VA’s agents which are not affected by any of the rules can either keep their current nonverbal behaviors or can be assigned any other. However, the system prevents contradictory nonverbal behaviors

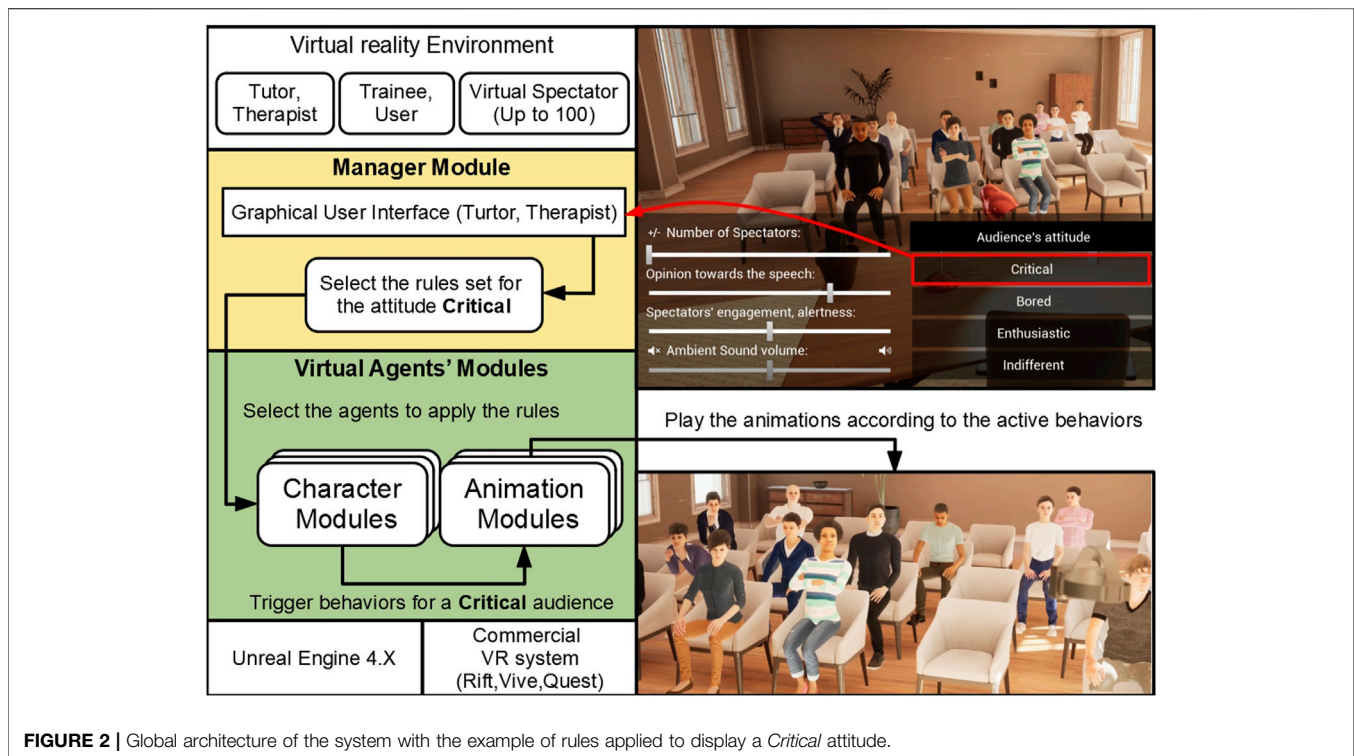
to be associated. It could lead to ambiguous agents behavior, e.g. to prevent agents to smile and shake the head at the same time or to nod while gazing away from the user. These sets of rules can also include more than one rule per categories and for instance including three different rules for the posture which would allow more complexity and variation in the agents’ behaviors. The Table 1 gives an example of rules set used to generate different attitudes.

### 3.2 System’s Structure

The system, provided as a plugin to the Unreal Engine, is composed of three main modules (Figure 2). The Manager module grants external control of the aforementioned nonverbal behavior rules, and selects and allocates rules for each individual virtual spectator composing the VA. In order to specify an audience, the user of the plugin can either select sets of pre-defined rules displaying a particular attitude or select and instantiate individual rules such as described in the previous section to specify custom behaviors. This module only provides a programming interface but can be paired with a 3D graphical user interface (GUI) which we developed in order to allow users to modify the VA in real time directly in VR. This allows non-programmers to easily design VAs. A VA’s attitude can thus also be tailored directly at runtime by instructors or therapists to elicit a particular user emotional response.

A Character module, one for each virtual spectator in the audience, receives the rules allocated by the Manager module and computes and triggers the animations’ blending for the targeted virtual spectator accordingly. Technically speaking, this module inherits from the UE4’s character which allows the combination of our system with all the advanced features available in the game engine, e.g., state machines, navigation, character’s senses modules. The Character module manages the behavior logic and computes the





**FIGURE 2** | Global architecture of the system with the example of rules applied to display a *Critical* attitude.

timing aspects of the various behaviors to trigger and blend according to the rules.

The actual animation blending is achieved through the Animation module, which computes the movements of the agent according to the nonverbal behaviors it has to display. The animation actually played out by an agent is dynamically constructed from one to several animations and may use rotations on characters' bones, each corresponding to a specific behavior type. There are at least three layers of animations in order to fully display an agent behavior: the posture, head, and gaze layers. The posture and the head layers, respectively responsible for the body and the head movements, are mixed so that these two first layers display a posture and a head movement at the same time, e.g., leaning backward, arms crossed and shaking the head. The gaze layer is responsible for moving the neck in a certain direction. There is a further optional layer which can be activated for the eye movement. The head layer and the gaze layer never blend together: this is to avoid any incoherent head movements. Facial expressions are handled separately with the use of morph targets (shape keys) which warp the face of the virtual spectator so that it can display the desired facial expression.

The plugin can handle 3D models and animations from a variety of free 3D Character Modeling tools such as Mixamo, or Autodesk Character Generator. The system can be easily enriched by adding new animations and head movements. It can also be enriched with new facial expressions by adding new morph targets to the rules. In the evaluation we describe here, we used seated agents. However, the system is flexible and allows for agents from different 3D Character Modeling tools, which can be placed in a variety of configurations and play stand-up animations. More details on the modules

implementation within the game engine and the combination and blending mechanism for the animations can be found in Glémarec et al. (2020) but without further information on our rules based approach which is a new contribution.

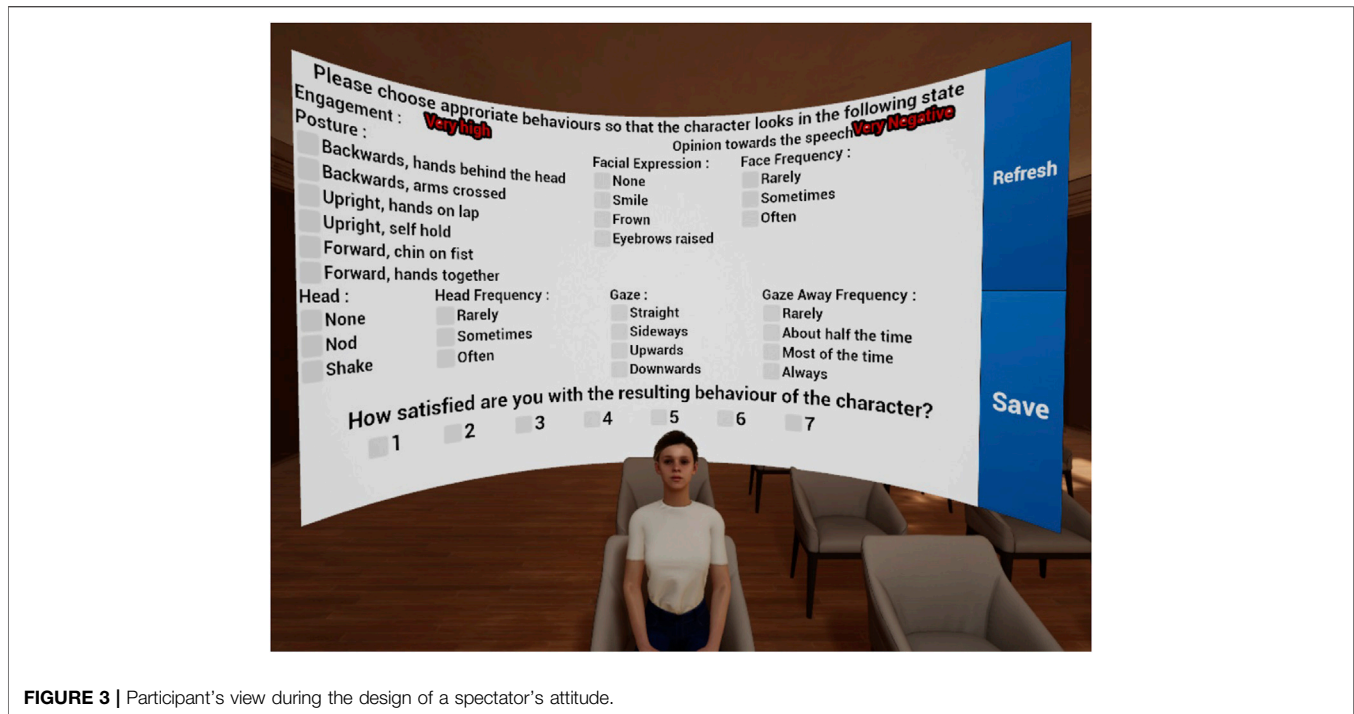
## 4 USER EVALUATIONS

### 4.1 First User Study: Nonverbal Behavior Perception of Individual Virtual Spectators

In this study, we investigated how nonverbal behaviors of individual agents are associated with the valence and the arousal dimensions. We aimed to confirm whether the use of these behaviors for different audience attitudes corresponded with a valence-arousal pair. Hence we asked participants to design the behavior of a spectator depending on a given pair of opinion and engagement towards a speech (for instance very low engagement and a negative opinion towards a speech).

#### 4.1.1 Hypothesis

During this study, we aimed to confirm the following hypotheses, which have been validated for desktop-based simulation by earlier work (Chollet and Scherer, 2017). The aim was to provide a set of validated nonverbal behavior for individual agents in order to build the audience model. In the following hypotheses, the independent variables (IV) are the levels of valence or arousal and the dependant variables (DV) are the different nonverbal behaviors the users can select from a GUI (Figure 3). Both IVs can take five different levels of valence or arousal, respectively very negative, negative, neutral, positive,



**FIGURE 3** | Participant's view during the design of a spectator's attitude.

very positive and very low, low, medium, high, very high. As for the DVs, we considered the same nonverbal behaviors as from the desktop study. Thus, the postures are evaluated in terms of proximity (leaning forward or backward) and openness (arms crossed or hand behind the head). For the rest of the behaviors visible on the **Figure 3**, we used the same behaviors as Chollet and Scherer (2017).

*Hypothesis 1 (H1.1)*, arousal and expressions: Higher arousal leads to more feedback, more facial expressions, more head movements, and more gaze directed at the speaker.

*Hypothesis 2 (H1.2)*, valence and expressions: Smiles and nods are associated with positive valence, frowns and head shakes with negative valence, and eyebrow raises and a face at rest with neutral valence.

*Hypothesis 3 (H1.3)*, arousal and postures: Postures chosen for high arousal involve leaning closer to the speaker than postures chosen for lower arousal.

*Hypothesis 4 (H1.4)*, valence and postures: Relaxed postures lead to a more positive valence compared with more closed postures.

#### 4.1.2 Method

For this study, 20 people participated, 4 women and 16 men, aged from 18 to 28, 13 of which are students and seven are in the workforce. Participants had to select behaviors for all the possible pairs of valence and arousal. Each value of valence and arousal has five levels: respectively very negative, negative, neutral, positive and very positive for the valence and very low, low, medium, high and very high for the arousal. Participants had to select the different behaviors on a GUI directly in VR (**Figure 3**). In doing so, the users could directly see the changes of behavior without

taking or removing the head mounted display. When satisfied by the resulting behavior, the participants could validate the answer and continue to the next pair of valence and arousal values. These pair values were randomized to avoid any order effect. The agents used were either male or female. The agent's gender was balanced in between each answer to avoid effects. The virtual environment was a simple room with chairs for the agent and a desk behind the user. Once the participant had finished specifying an agent's behavior, they were asked to rate how satisfied they were with the result using a 7-point satisfaction scale. A poor rating would have led to the behavior being discarded from the model. However none of the participants gave a rating under four for any of the agents. Each session started with a short training phase where the users were able to become familiar with the virtual environment and the GUI. There was no time limit set for designing each agent. At the end of the session, participants were interviewed about their overall experience. To carry out this study we used a HTC®Vive Pro running on a stationary computer running with Windows 10 64 bits, Intel®Core i7-7700k processor (8 cores, 4.20 GHz, 11 MB cache, 11 GT/s) and NVIDIA®GeForce GTX 1080Ti Graphics Processing Unit (GPU) 16 GB GDDR5. The simulation runs on average with 90 frame per second on this computer.

#### 4.1.3 Results

For H1.1, H1.3, and H1.4 we use non-binary factors and ordinal numerical variables. The postures which were described by openness and the proximity transformed into numerical variables where a high proximity value represents a forward posture and low represents a backward posture (backward is 1 and forward is 3), and a high openness value represents open and a low value represents closed (arms crossed and self-hold: 1, arms

**TABLE 2** | Wilcoxon's Pairwise Tests, numbers are  $p$ -values from each tests, the levels of arousal are on both sides of the table with the Spearman's effect size for the arousal and the four variables.

<b>Gaze away frequency</b>					
Arousal	High	Low	Medium	Very high	Spearman's effect size ( $\rho$ )
Low	<0.01	-	-	-	-0.6
Medium	<0.01	<0.01	-	-	
Very high	0.462	<0.01	<0.01	-	
Very low	<0.01	0.054	<0.01	<0.01	
<b>Facial expressions frequency</b>					
Arousal	High	Low	Medium	Very high	Spearman's effect size ( $\rho$ )
Low	0.012	-	-	-	0.2
Medium	0.345	1.0	-	-	
Very high	1.0	0.003	0.689	-	
Very low	0.040	1.0	1.0	0.016	
<b>Head movements frequency</b>					
Arousal	High	Low	Medium	Very high	Spearman's effect size ( $\rho$ )
Low	<0.01	-	-	-	0.3
Medium	0.042	1.0	-	-	
Very high	1.0	<0.01	<0.01	-	
Very low	<0.01	1.0	0.560	<0.01	
<b>Proximity</b>					
Arousal	High	Low	Medium	Very high	Spearman's effect size ( $\rho$ )
Low	<0.01	-	-	-	0.7
Medium	<0.01	<0.01	-	-	
Very high	1.0	<0.01	<0.01	-	
Very low	<0.01	0.079	<0.01	<0.01	

behind the head: 3, the rest: 2). We used the exact same transformation as in Chollet and Scherer (2017). We also transformed the frequencies into an ordinal variable similarly to the model. Where no behavior is 0, *rarely* is 0.25, *sometimes* and *about half the time* are 0.5, *often* and *most of the time* are 0.75, and *Always* is 1. In doing so, we can easily compare our results to those obtained from the literature. It is also necessary because we used a within-subjects design and non-binary nominal factors which prevent us from using Pearson's Chi squared test or non-parametric Cohen's test for independence. Hence, because the distribution is not normal, we ran a non-parametric Friedman test instead of a repeated measures ANOVA test.

#### 4.1.3.1 Arousal and Expressions

For H1.1, we set the arousal as the IV and conducted tests with the face, head and gaze-away frequencies as DV. Pairwise Wilcoxon's tests using Bonferroni's adjustment method have been used on each feature of the IV. For all three DVs, the arousal has a significant effect on the behaviors' frequencies (gaze:  $\chi^2 = 192, df = 4, p\text{-value} < 0.05$ , facial expressions:  $\chi^2 = 17.7, df = 4, p\text{-value} < 0.05$ ), head movements:  $\chi^2 = 47.6, df = 4, p\text{-value} < 0.05$ ). The pairwise tests (Table 2) show which levels are significant compared to the others. With regard to the facial expressions, low arousal leads to less frequent facial expression and a high arousal to more frequent facial expressions. The participants could not significantly differentiate two levels of arousal which were not markedly different, for instance *Very Low* and *Low*. The Figure 4

gives a representation of these differences while showing which levels are significant to each other. However it is important to underline that the effect size for the facial expression frequency and the arousal is small (Spearman's  $\rho = 0.2$ ). Thus we can only slightly agree with these conclusions.

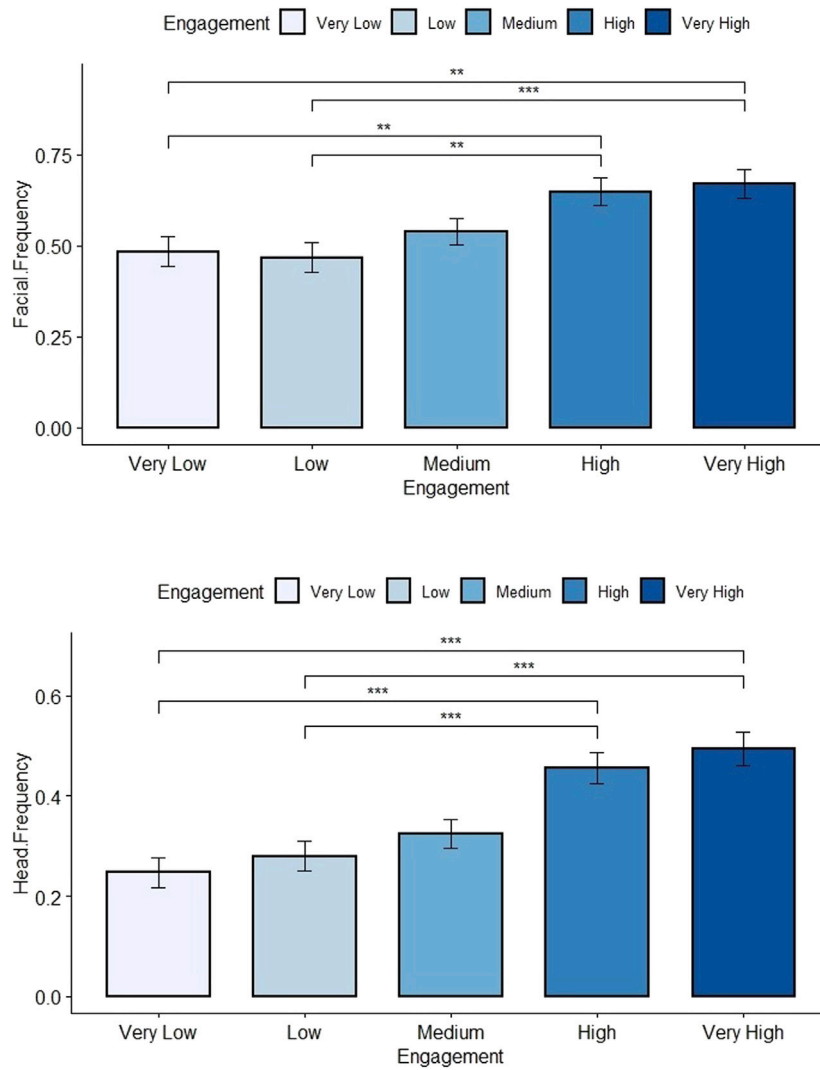
The test results for the head movement frequency are much alike. A low arousal leads to less frequent head movement, and a high arousal leads to more frequent head movements. The compared frequencies are only significant for opposed levels of arousal, e.g., *Low* and *High* arousal (Figure 4). For this two first results, the behavior frequency for a medium arousal is not significantly different from all other levels. This is probably due to the size of the frequency scale, which only had three levels. The following tests had four levels, which then allowed the users to design behaviour with significant different behaviour frequencies.

Therefore, the gaze-away frequency is significant for all levels of arousal except between a *High* and a *Very High* level and a *Low* and a *Very Low* level of arousal. This means that a higher arousal level leads to less frequent gaze-away while a low level of arousal leads to more frequent gaze-away from the user (Figure 5). Overall, a lower arousal leads to less frequent expressions and a high arousal leads to more frequent expressions.

#### 4.1.3.2 Arousal and Postures

For H1.3, we set the arousal as the IV and the proximity as the ordinal DV. Results are significant as well ( $\chi^2 = 221, df = 4, p\text{-value} < 0.05$ ); higher arousal leads to a





**FIGURE 4** | Distribution plots of the facial expressions and head movements frequency depending on the engagement (Arousal) levels. Significant results with the *Medium Engagement* level was removed for clarity purpose, **Table 2** for details.

closer posture proximity, while a lower arousal leads to a further posture. The pairwise tests show that the proximity is mainly significant for markedly different levels of arousal: for instance for a *High* and a *Low* level of arousal. The **Table 2** and **Figure 5** resume these results.

**4.1.3.3 Valence and Postures**

For H1.4, we set the valence as the IV and relaxation as the DV. Surprisingly, this test is not significant so we failed to reject the null hypothesis. In our case, the relaxation or the openness of the proposed postures does not seem to be related to the valence. This result differs from the findings in the desktop-based audience research by Chollet and Scherer (2017).

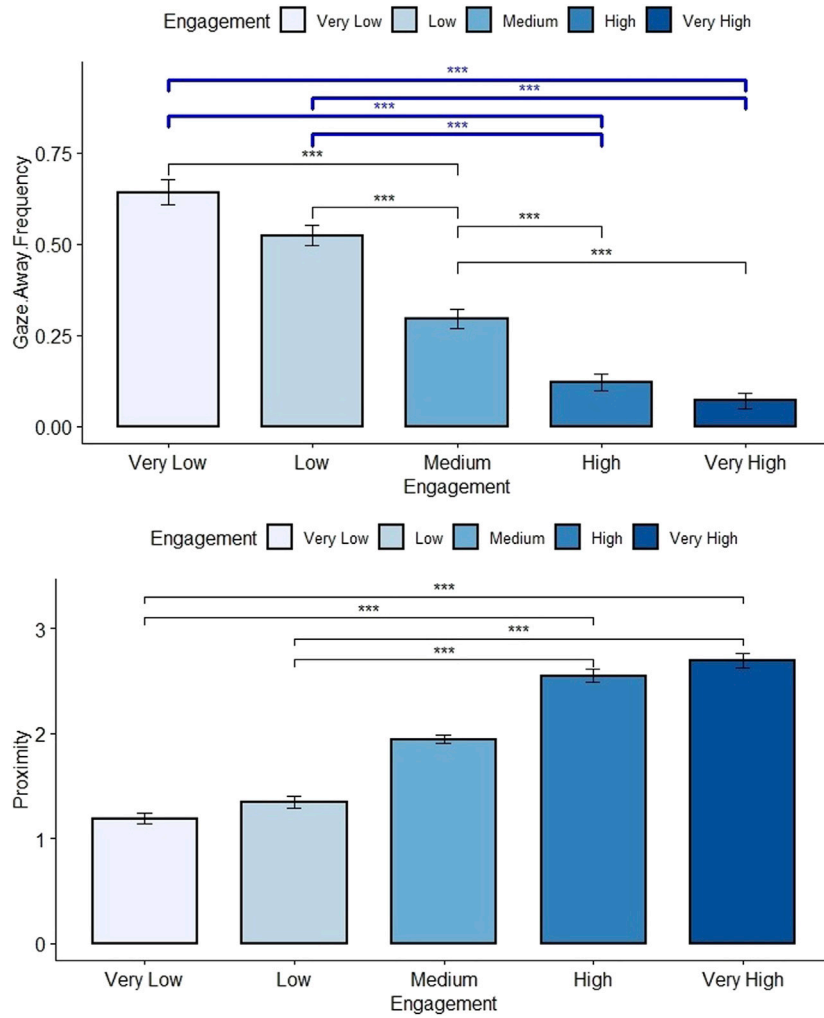
**4.1.3.4 Valence and Expressions**

Finally, for H1.2, we set the valence as the IV and conducted tests with the facial expression and head movement categories as the

nominal DVs. Here, both the IV and DVs have more than two levels and the study was within-subjects design, so we used a multinomial logistic regression. If we do not transform our DV into ordinal data, it is because we are interested in getting the influence of each behavior type per level of valence, unlike the previous tests in which we were comparing mean frequencies per subject or the average proximity values for the posture. Hence we used a face without facial expressions for the agent as the reference event to determine all odds ratios for the IV. The regression model is expressed as:

$$g_j = \beta_0 + \beta_j * x_i \tag{3}$$

where  $x_i$  represents the different levels of valence,  $j$  the behavior parameter selected by the participants, and  $\beta_0$  is the intercept parameter. Results regarding the head movements indicate that shaking the head is associated with a negative valence, and



**FIGURE 5 |** Distribution plot of the gaze away frequency and the posture proximity depending on the engagement (Arousal) levels. Significant results with the *Medium Engagement* level was removed for clarity purpose. Main significant results from the pairwise tests related to the gaze are shown in blue above the bar charts, **Table 2** for details.

**TABLE 3 |** Multinomial Logistic Regression significance table.

H1.2 head mov	Coeff	Std. error	Zvalue	Pr(> z )
Nod:(intercept)	-1.9	0.3	-6.3	<0.05
Opinion positive	2.9	0.4	7.7	<0.01
Opinion very positive	3.4	0.4	8.5	<0.01
Shake:(intercept)	-4.5	1.0	-4.4	<0.05
Opinion negative	5.7	1.0	5.5	<0.01
Opinion very negative	5.9	1.0	5.6	<0.01

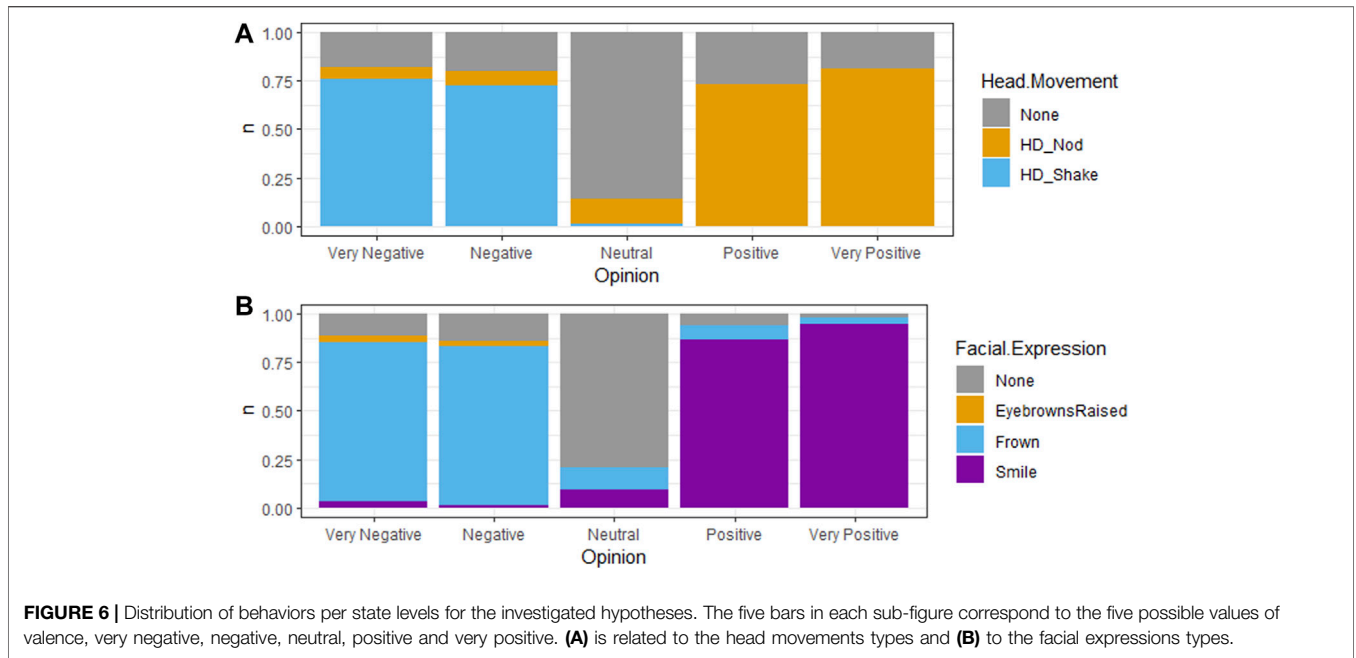
H1.2 facial exp	Coeff	Std. error	Zvalue	Pr(> z )
Frown:(intercept)	-1.9	0.3	-6.0	<0.01
Opinion negative	3.7	0.4	8.5	<0.01
Opinion very negative	3.8	0.4	8.6	<0.01
Smile:(intercept)	-2.2	0.4	-6.1	<0.01
Opinion positive	4.8	0.5	8.8	<0.01
Opinion very positive	6.0	0.8	7.5	<0.01

nodding with a positive valence (with an odds ratio within the confidence interval CI95%). Head shake is significantly negative and head nod is significantly positive. Details on the tested features are visible in **Table 3** and **Figure 6** shows the distribution for each behaviours depending on the levels of valence.

$$g_{shaking} = -4.5 + 5.7 * x_{negative} - 9.8 * x_{positive} + 5.9 * x_{veryNegative} - 9.6 * x_{veryPositive} \tag{4}$$

$$g_{nodding} = -1.9 + 0.9 * x_{negative} + 2.9 * x_{positive} + 0.8 * x_{veryNegative} + 3.3 * x_{veryPositive} \tag{5}$$

In order to study the relationship between the different facial expressions and the valence, we used the same formula (**Eq. 3**)



where  $j$  represents the facial expression selected by the participants. Results indicate that smiling is significantly associated with positive valence, and frowning with negative valence (with an odds ratio within the confidence interval CI95%). Almost none of the participants chose eyebrows raised, and this is why we do not propose an analysis for it. The **Figure 6** highlights the lack of selections of eyebrows raised.

$$g_{smile} = -2.8 - 0.4 * x_{negative} + 4.8 * x_{positive} + 0.9 * x_{veryNegative} + 6.0 * x_{veryPositive} \tag{6}$$

$$g_{frown} = -1.9 + 3.6 * x_{negative} + 2.0 * x_{positive} + 3.84 * x_{veryNegative} + 2.2 * x_{veryPositive} \tag{7}$$

These results partially confirm the findings from Chollet and Scherer (2017) with an exception to the *eyebrows raised* behavior with which we have no available data. We can also had that the default face displayed by the virtual agents was significantly preferred to the other facial expressions; this confirm our hypothesis on the association with a neutral valence and a face at rest.

### 4.2 Second User Study: Virtual Audience Attitudes Perception Evaluation

This study consisted of an evaluation aimed to validate the perceived audience attitudes generated by our model. Based on the nonverbal behavior rules using values of valence and arousal and the results from our first user evaluation, we designed different VAs. We investigated whether the audience attitudes generated with these rules could be identified by the users in terms of valence and arousal. The relationships between the nonverbal behaviors and the valence or the arousal are mainly from Chollet and Scherer (2017)’s studies in which they provide the proportion of agents with a certain

behavior needed to let the users recognize the audience’s attitude. The section below describes how we used the literature and our first study to build the VAs used in this study.

#### 4.2.1 Hypothesis

Our hypotheses for this study are that VR users can significantly perceive different attitudes generated with our system in terms of valence and arousal in VR. This means the model we use to generate the VA attitudes can be used to create the virtual agents’ behaviors which allow the users to perceive the targeted attitude.

*Hypothesis 1 (H2.1):* The higher the positive valence, the higher the positive perceived opinion is for the participant (respectively for negative valence and opinion).

*Hypothesis 2 (H2.2):* The lower the arousal, the lower the perceived engagement is for the participant (respectively for high arousal and perceived engagement).

In the above two hypotheses, the terms *high valence* and *low arousal* express the intensity of the displayed audience’s attitude. Hence, in this study, the VAs were populated with a higher proportion of virtual spectators displaying the behaviors matching with the targeted audience attitude. For instance, an attitude with a very positive valence and a very low arousal corresponds to an audience where more than 60% of the agents display a positive behaviour and where more than 30% of them are highly engaged (Chollet and Scherer, 2017). The behaviors from the first study are then used to generate the VA. For this example, what we call an agent exhibiting a positive behavior is an agent displaying the nonverbal behaviours corresponding to a positive valence (head nod and smile). The engagement for this same agent is also based on the behavior frequencies previously shown in our first study where a highly engaged agent would frequently nod and smile while leaning



**FIGURE 7** | Participant's view during the audience perception evaluation.



**FIGURE 8** | Example of an annoyed (A) and an interested (B) virtual audiences.

forward. It is also to consider that according to Chollet and Scherer (2017), the more frequent a behavior is, the stronger it will be perceived in the overall audience's attitude. The **Table 1** gives an example of what parameters we use to populate our VAs for different attitudes. Finally, for the attitude displaying a neutral valence and a medium arousal, we used a mix of positive, negative and neutral agents as well as a mix of agents with a low, medium or high engagement like in Chollet and Scherer (2017).

#### 4.2.2 Method

For this study, we kept the same virtual environment as that we had for the first evaluation. We used 10 different characters from Adobe® Mixamo, 5 females and 5 males (**Figure 7**). All these virtual spectators provided implementation of facial expressions. They were driven by our system according to the nonverbal behaviors previously identified during the individual spectators nonverbal behavior study (**Section 4.1**). **Figure 8** provides examples of both a critical and an interested audience.

Thirty eight people participated in the evaluation: 9 women and 29 men aged from 19 to 28. None of them participated in the first study and all were students. Participants had to rate their perceived audience's opinion and engagement on two 5-point scales representing the value of valence and the level of arousal (*i.e.* from 1 to 5: very negative, negative, neutral, positive and very

positive valence and very low, low, medium, high and very high for the arousal). The same audiences were shown to participants but in randomized orders to avoid any effects. These audiences were designed to correspond to five different types of attitudes:

- Attitude 1, (A1):* very negative valence and very low arousal;
- Attitude 2, (A2):* very negative valence and very high arousal;
- Attitude 3, (A3):* neutral valence and medium arousal;
- Attitude 4, (A4):* very positive valence and very low arousal;
- Attitude 5, (A5):* very positive valence and very high arousal.

Each session started with a short training phase where the users were able to become familiar with the virtual environment and the GUI. There was no time limit set for giving an answer. At the end of the session, users were interviewed about their overall experience. The same hardware was used in this evaluations.

#### 4.2.3 Results

The evaluation followed a within-subjects design, and the distribution was not normal, so we ran a non-parametric Friedman test for H2.1 and H2.2. With regard to H2.1, we set the generated attitudes defined above ( $A_i$ ) as the IV and conducted a test with the perceived opinion value as the DV. For H2.2, we also set the generated attitudes defined above ( $A_i$ ) as

**TABLE 4 |** Wilcoxon's Pairwise Tests, numbers are  $p$ -values from each tests, the attitudes are on both sides of the table with the Spearman's effect size for the valence and then the arousal.

Valence					
Attitudes	Very low and very negative	Very high and very negative	Medium and neutral	Very low and very positive	Spearman's effect size ( $\rho$ )
Very high and very negative	1.0	-	-	-	
Medium and neutral	<0.01	<0.01	-	-	0.5
Very low and very positive	<0.01	<0.01	1.0	-	
Very high and very positive	<0.01	<0.01	1.0	1.0	
Arousal					
Attitudes	Very low and very negative	Very high and very negative	Medium and neutral	Very low and very positive	Spearman's effect size ( $\rho$ )
Very high and very negative	1.0	-	-	-	
Medium and neutral	1.0	1.0	-	-	0.2
Very low and very positive	<0.01	<0.01	<0.01	-	
Very high and very positive	1.0	1.0	1.0	<0.01	

the IV and conducted a test with the perceived engagement value as the DV. Pairwise Wilcoxon's tests using the Bonferroni adjustment method have been used for each modality (Table 4).

For both the valence and the arousal values, we found a significant effect on the perceived attitudes (valence:  $\chi^2 = 43.7$ ,  $df = 4$ ,  $p$ -value < 0.01, arousal:  $\chi^2 = 26.4$ ,  $df = 4$ ,  $p$ -value < 0.01). In the pairwise test for H2.1, the three different values of valence are correctly associated with the audience attitudes by the participants (negative, neutral and positive). Moreover the perceived values of valence are only significant when comparing between attitudes from markedly different levels of valence: there is no differences between attitudes with the same value of valence (Figure 9). However, the neutral audience (A3) is not significantly perceived as neutral by the participants but slightly positive. We believe this is due to the fact that we mixed different type of nonverbal behaviors which where not only neutral.

For H2.2, the pairwise test shows that participants cannot significantly identify all levels of arousal. The results exhibit a significant difference of the users' perception between low and high arousal when the valence is positive (A4 and A5, Figure 9). However, in A1, A2, and A3, there are no significant differences in terms of perceived arousal. Participants cannot significantly differentiate two different levels of arousal when the valence is negative. The same for A3 which is supposed to be perceived as moderately engaged was perceived as highly engaged. We believe it is also due to the use of negative nonverbal behaviors to generate the attitude. All three attitudes we designed with negative nonverbal behaviors were perceived with the same high level of arousal (Figure 9). Further investigations can be done to test if negative nonverbal behaviors also influence the perceived arousal. Thus, H2.2 is only partially validated, and we cannot completely reject the null hypothesis. An example of attitudes the users cannot significantly recognize are reported on the Figure 1.

### 4.3 Recommendations for Audience Simulation in Virtual Reality

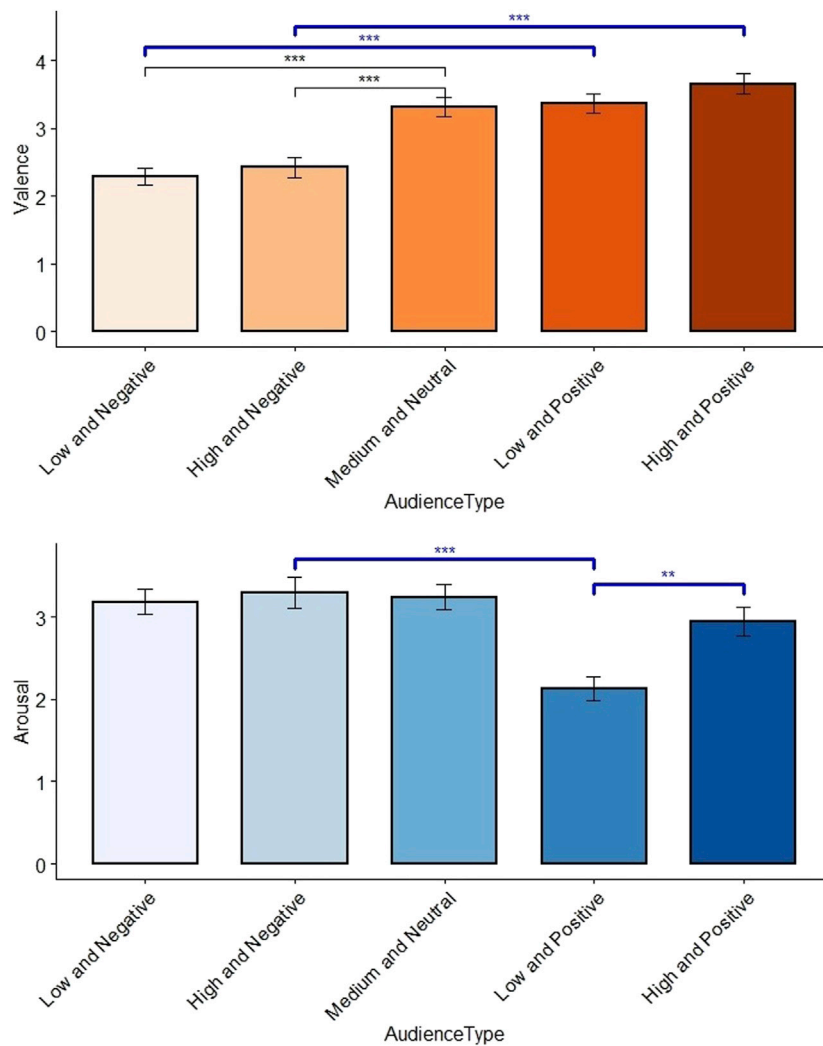
With respect to our results and considering the existing literature, we can propose recommendations and guidelines for the generation of different attitudes and the design of Virtual Audiences. These guidelines are similar to those we used to build our system and only concern nonverbal behaviors. Other elements such as the sound or backchannels are discussed in the following section.

**Guideline, Engagement towards the speech (Arousal):** the engagement of an audience is significantly related to the gaze, the frequency of movements and by the posture's proximity. However, for facial expressions, we would advise to alternate between the targeted facial expression and a face at rest in order to allow the users to perceive and compare those differences (Kang et al., 2016) instead of displaying them continuously.

**Guideline, Opinion towards the speech (Valence):** the opinion is significantly related to the nonverbal behavior type. Thus, we would advise using very distinct head movements and facial expressions for spectators, which are clearly identified by the users. Subtle facial expressions and head movements might be perceived if a VR user is close to the agent but proportionately less if distant to it. These two behaviors seem to be more easily characterised in terms of opinion when compared to the posture openness. Finally, it appears that negative nonverbal behaviors such as frowning or shaking the head are more perceptible than positive ones like smiling or nodding.

**Recommendation, negative audiences and engagement:** according to our results, users may have difficulties differentiating levels of engagement (Arousal) when the VA has a negative opinion (Valence). Consequently, we would recommend using a lower percentage of nonverbal behaviors recognised as a display of a negative opinion (such as frowning or shaking the head and crossing the arms) when designing audiences with a negative attitudes as compared to the percentage of positive non verbal





**FIGURE 9** | Distribution plot of the selected valence and arousal depending on the attitudes. Main significant results from the pairwise tests for our hypotheses are shown in blue above the bars charts.

behavior used to simulate positive audiences. This is further supported by Chollet and Scherer (2017) which reported that fewer of these behaviors are required to let the users correctly recognize a negative audience. Users might then be able to better recognize behaviors associated with the engagement like the gaze. With regard to our results, negative nonverbal behaviors seem to increase the perceived level of engagement.

## 5 DISCUSSION AND LIMITATIONS

Our plugin offers a high-level building tool for audiences with varying attitudes. The implementation described in Section 3 offers the VR community the opportunity to design and control VAs with precise parameters. VR GUIs, as shown in Figure 3, offer the possibility of easy design of the VA. These interfaces also facilitate direct in-game manipulation of the audiences. A variety of applications of VAs, such as public speaking anxiety exposure

therapy (Harris et al., 2002), classroom management (Lugrin et al., 2016), and social skills training (Chollet et al., 2018) will be able to take advantage of our interface to create audiences with specific, low-level behaviors corresponding to particular unique attitudes.

In our future work, we plan to further develop the audience authoring by providing higher-level user control to manipulate the audiences: this could benefit VR training systems, with regard to the trade-off between a fully autonomous simulation and a Wizard of Oz system where each spectator would be individually controlled. For instance, when replacing tutor expertise with an autonomous component is not desirable, e.g., VR therapy and training could need real time adjustments and a fine control of the environment.

Our system could then bridge this gap and provide a component which may be easily used by the tutor during a simulation-based training session. Additionally, these VR applications could benefit from a more immersive and interactive VA. Nevertheless, even if the VAs' attitudes are



**FIGURE 10** | Example of nonverbal behaviors used to display four different attitudes, **(A)** shows this behaviors on one agent and **(B)** for the entire virtual audience.

dynamic, there is still some space for improvement, whether in the behavior model or in the potential interactions with the users.

We plan to engage domain experts, such as therapists for VR exposure therapy or classroom management experts for VR teacher training, in order to add domain-specific circumstances and knowledge to our VAs. Finally, we intend to investigate the interactions between users and the VAs by studying users' proximity to the VAs and gaze. Thus, we will investigate the impact of a speaker's location on the audience's behavior but also the effect of the user's gaze on a specific agent from the VA. Such interactions will enhance the simulation's believability and, as a result, improve the feeling of social-presence and immersion.

Our evaluations have shown that despite not being domain-specific, we can successfully generate different kinds of audiences that are significantly identified by users such as *indifferent*, *critical*, and *enthusiastic* (Figure 10). We broadly managed to validate our hypotheses on VAs' attitude perception and confirm its efficiency in virtual reality, with the exception of VAs displaying low engagement and negative opinion as per the model. Low engagement and negative opinion are not

identified correctly by VR users, which might prevent the perception of bored or anxious attitudes. It probably comes from the fact that negative behaviors are more easily recognized by VR users than those related to the arousal.

Our results also raise questions about some behaviors from the model and their perceptions. Posture relaxation, deemed significant for the perception of valence in the model, has not been confirmed as such in our first study. However, users still significantly distinguish a positive audience from a negative one in our second study when the posture relaxation is used. It probably comes from the fact that negative behaviors are more easily recognized by VR users than those related to the arousal.

The interviews with the participants gave us a more insight into their understanding and perception of the audiences' attitudes. Some of them mentioned that the facial expressions were a very strong signal, which corresponds to the results from our first evaluation. Participants also mentioned the fact that they were using a known context in order to make their decision, *e.g.*, a lecture or a professional meeting. No such context was given before the experiment, but a precise context might help the users to

associate nonverbal behaviors with a known audience attitude. Lastly the most recurrent comment was the lack of sounds when some behaviors were played. None of the studies we used to build our model define sound backchannels, and thus we avoided sound in our evaluations as well. However, other studies concerning virtual agent behavior, such as research of conversational agents, may fill in the gap between the nonverbal behaviors and the associated backchannels (Laslett and Smith, 2002; Poppe et al., 2010; Kistler et al., 2012; Barmaki and Hughes, 2018).

Based on this feedback, we can see that some of the limitations in the audience's attitude perception may be bypassed when using the system in a specific context. Because more behavior diversity and sound may also help to improve perception, some of these features are already being developed as part of the system such as sound backchannels, shared behaviors like spectator chatting together but also movements. Virtual spectators have the ability to move in the audience and display stand-up animations as well. These behaviors were not part of the evaluations we present here in order to preserve the integrity of the relationships we tested. However, behaviours such as chatting, moving, etc. warrant exploration in future evaluations.

## 6 CONCLUSION

In this paper, we described the design and evaluations of a virtual audience simulation system for VR applications. This system implements a behavior model defining a set of nonverbal behavior rules. Each rule specifies a nonverbal behavior category: *posture*, *head movement*, *facial expression* and *gaze direction* as well as three parameters: *type*, *frequency* and *proportion*. In a first user-study, we asked participants to create sets of nonverbal behaviour parameters to simulate different attitudes. We then used these parameters in a second user-study to design and evaluate different types virtual audiences attitudes. Overall, our results confirmed the system's capacity to simulate virtual audiences with three types of different perceived attitudes: *indifferent*, *critical*, and *enthusiastic*. In addition, we also proposed a set of recommendations and guidelines regarding attitudes and expressions for future design of audiences for VR therapy and training applications. Our future work is two-fold. First, we will improve the overall quality and variety of the virtual spectator animations as well as provide automatic audience

reactions to the speech and nonverbal behaviours of a user in VR using nonverbal cues like eye contact, prolonged gaze, and proximity (Laslett and Smith, 2002; Kistler et al., 2012; Barmaki and Hughes, 2018). Secondly, we will integrate and evaluate our system within a VR therapy application dealing with public speaking anxiety by focusing on the treatment of elocution problems such as stuttering.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: <http://hci.uni-wuerzburg.de/projects/virtual-audiences/>

## ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

YG, A-GB, and J-LL contributed to conception and design of the study. YG developed the software and performed the statistical analyses. AJ and YG performed the user evaluation. A-GB, J-LL, CB, and ML were the project administrators and supervised the work. All authors contributed to manuscript revision, read, and approved the submitted version.

## FUNDING

YG is funded by the ENIB, French Ministry of Higher Education, Research and Innovation. This work is PARTIALLY funded by the VIRTUALTIMES project (ID-824128) funded by the European Union under the Horizon 2020 program. This publication was supported by the Open Access Publication Fund of the University of Wuerzburg.

## REFERENCES

- Anderson, P. L., Price, M., Edwards, S. M., Obasaju, M. A., Schmertz, S. K., Zimand, E., et al. (2013). Virtual Reality Exposure Therapy for Social Anxiety Disorder: A Randomized Controlled Trial. *J. consulting Clin. Psychol.* 81, 751–760. doi:10.1037/a0033559
- Anderson, P. L., Zimand, E., Hodges, L. F., and Rothbaum, B. O. (2005). Cognitive Behavioral Therapy for Public-Speaking Anxiety Using Virtual Reality for Exposure. *Depress. Anxiety* 22, 156–158. doi:10.1002/da.20090
- Arthur, K. W., Booth, K. S., and Ware, C. (1993). Evaluating 3d Task Performance for Fish Tank Virtual Worlds. *ACM Trans. Inf. Syst.* 11, 239–265. doi:10.1145/159161.155359
- Barmaki, R., and Hughes, C. (2018). Gesturing and Embodiment in Teaching: Investigating the Nonverbal Behavior of Teachers in a Virtual Rehearsal Environment. *Proc. AAAI Conf. Artif. Intelligence* 32.

- 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* (Springer), 116–128. doi:10.1007/978-3-642-40415-3\_10
- Biocca, F., Harms, C., and Burgoon, J. K. (2003). Toward a More Robust Theory and Measure of Social Presence: Review and Suggested Criteria. *Presence: Teleoperators & Virtual Environments* 12, 456–480. doi:10.1162/105474603322761270
- Bowman, D., Kruijff, E., LaViola, J. J., Jr, and Poupyrev, I. P. (2004). *3D User Interfaces: Theory and Practice*. CourseSmart eTextbook (Addison-Wesley).
- Chollet, M., Ghate, P., and Scherer, S. (2018). "A Generic Platform for Training Social Skills with Adaptive Virtual Agents, in Proceedings Of the 17th International Conference On Autonomous Agents And MultiAgent Systems (Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems) (AAMAS), '18, 1800–1802.

- Chollet, M., and Scherer, S. (2017). Perception of Virtual Audiences. *IEEE Comput. Grap. Appl.* 37, 50–59. doi:10.1109/mcg.2017.3271465
- Chollet, M., Sratou, G., Shapiro, A., Morency, L.-P., and Scherer, S. (2014). “An Interactive Virtual Audience Platform for Public Speaking Training,” in *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems*. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems), 1657–1658.
- Cummings, J. J., and Bailenson, J. N. (2016). How Immersive Is Enough? a Meta-Analysis of the Effect of Immersive Technology on User Presence. *Media Psychol.* 19, 272–309. doi:10.1080/15213269.2015.1015740
- Delamarre, A., Lisetti, C., and Buche, C. (2020). “Modeling Emotions for Training in Immersive Simulations (Metis): A Cross-Platform Virtual Classroom Study,” in *2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct*. Los Alamitos, CA: ISMAR-Adjunct, 78–83. doi:10.1109/ISMAR-Adjunct51615.2020.00036
- Ekman, P. (1999). Basic Emotions. *Handbook Cogn. Emot.* 98, 16. doi:10.1002/0470013494.ch3
- Fukuda, M., Huang, H.-H., Ohta, N., and Kuwabara, K. (2017). “Proposal of a Parameterized Atmosphere Generation Model in a Virtual Classroom,” in *Proceedings of the 5th International Conference on Human Agent* (New York, NY: Interaction (ACM)), 11–16.
- Gebhard, P., Schneeberger, T., Baur, T., and André, E. (2018). “Marssi: Model of Appraisal, Regulation, and Social Signal Interpretation,” in *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, 497–506.
- Glémarec, Y., Lugrin, J., Bossier, A., Cagniat, P., Buche, C., and Latoschik, M. E. (2020). “Pushing Out the Classroom walls: A Scalability Benchmark for a Virtual Audience Behaviour Model in Virtual Reality,” in *Mensch und Computer 2020 - Workshopband, Magdeburg, Germany, September 6-9, 2020*. Editors C. Hansen, A. Nürnberger, and B. Preim (Gesellschaft für Informatik e.V). doi:10.18420/muc2020-ws134-337
- Guimarães, M., Prada, R., Santos, P. A., Dias, J., Jhala, A., and Mascarenhas, S. (2020). “The Impact of Virtual Reality in the Social Presence of a Virtual Agent,” in *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents* (New York, NY, USA: Association for Computing Machinery), 1–8. IVA '20. doi:10.1145/3383652.3423879
- Hale, K. S., and Stanney, K. M. (2006). Effects of Low Stereo Acuity on Performance, Presence and Sickness within a Virtual Environment. *Appl. Ergon.* 37, 329–339. doi:10.1016/j.apergo.2005.06.009
- Harris, S. R., Kemmerling, R. L., and North, M. M. (2002). Brief Virtual Reality Therapy for Public Speaking Anxiety. *CyberPsychology Behav.* 5, 543–550. doi:10.1089/109493102321018187
- Hayes, A. T., Hardin, S. E., and Hughes, C. E. (2013). *Virtual, Augmented and Mixed Reality. Systems and Applications: 5th International Conference, VAMR 2013, Held as Part of HCI International 2013, Las Vegas, NV, USA, July 21–26, 2013, Proceedings, Part II*. Berlin, Heidelberg: Springer Berlin Heidelberg), 142–151. chap. Perceived Presence’s Role on Learning Outcomes in a Mixed Reality Classroom of Simulated Students. doi:10.1007/978-3-642-39420-1\_16
- Heudin, J.-C. (2004). “Evolutionary Virtual Agent,” in *Proceedings. IEEE/WIC/ACM International Conference on Intelligent Agent Technology*, 2004. (IAT 2004) (Los Alamitos, CA: IEEE), 93–98.
- Heudin, J.-C. (2007). “Evolutionary Virtual Agent at an Exhibition,” in *International Conference on Virtual Systems and Multimedia* (Berlin, Heidelberg: Springer), 154–165.
- Hosseinpah, A., Krämer, N. C., and Straßmann, C. (2018). “Empathy for Everyone? the Effect of Age when Evaluating a Virtual Agent,” in *Proceedings of the 6th International Conference on Human-Agent Interaction*, 184–190.
- Kahlon, S., Lindner, P., and Nordgreen, T. (2019). Virtual Reality Exposure Therapy for Adolescents with Fear of Public Speaking: a Non-randomized Feasibility and Pilot Study. *Child. Adolesc. Psychiatry Ment. Health* 13, 1–10. doi:10.1186/s13034-019-0307-y
- Kang, N., Brinkman, W.-P., Birna van Riemsdijk, M., and Neerinx, M. (2016). The Design of Virtual Audiences: Noticeable and Recognizable Behavioral Styles. *Comput. Hum. Behav.* 55, 680–694. doi:10.1016/j.chb.2015.10.008
- Kistler, F., Endrass, B., Damian, I., Dang, C. T., and André, E. (2012). Natural Interaction with Culturally Adaptive Virtual Characters. *J. Multimodal User Inter.* 6, 39–47. doi:10.1007/s12193-011-0087-z
- Laslett, R., and Smith, C. (2002). *Effective Classroom Management: A Teacher’s Guide*. New York, NY: Routledge. doi:10.4324/9780203130087
- Latoschik, M. E., Kern, F., Stauffert, J.-P., Bartl, A., Botsch, M., and Lugrin, J.-L. (2019). Not alone Here?! Scalability and User Experience of Embodied Ambient Crowds in Distributed Social Virtual Reality. *IEEE Trans. Vis. Comput. Graphics* 25, 2134–2144. doi:10.1109/vtcg.2019.2899250
- Lee, C., Bonebrake, S., Bowman, D. A., and Höllerer, T. (2010). “The Role of Latency in the Validity of Ar Simulation,” in *2010 IEEE Virtual Reality Conference (VR) (IEEE)*, 11–18.
- Lugrin, J.-L., Latoschik, M. E., Habel, M., Roth, D., Seufert, C., and Grafe, S. (2016). Breaking Bad Behaviours: A New Tool for Learning Classroom Management Using Virtual Reality. *Front. ICT* 3, 26. doi:10.3389/fict.2016.00026
- Lugrin, J.-L., Wiebusch, D., Latoschik, M. E., and Strehler, A. (2013). “Usability Benchmarks for Motion Tracking Systems,” in *Proceedings Of the 19th ACM Symposium On Virtual Reality Software And Technology (ACM)VRST*, 13, 49–58.
- Marsella, S., and Gratch, J. (2002). “A Step toward Irrationality: Using Emotion to Change Belief,” in *Proceedings Of the First International Joint Conference on Autonomous Agents and Multiagent Systems: Part 1*, 334–341.
- Mehrabian, A. (1996). Pleasure-arousal-dominance: A General Framework for Describing and Measuring Individual Differences in Temperament. *Curr. Psychol.* 14, 261–292. doi:10.1007/bf02686918
- Messinis, I., Saltouras, D., Pintelas, P., and Mikropoulos, T. (2010). “Investigation of the Relation between Interaction and Sense of Presence in Educational Virtual Environments,” in *2010 International Conference on E-Education (e-Business, e-Management and e-Learning)*, 428–431. doi:10.1109/ICAE.2010.137
- Narayan, M., Waugh, L., Zhang, X., Bafna, P., and Bowman, D. (2005). “Quantifying the Benefits of Immersion for Collaboration in Virtual Environments,” in *Proceedings of the ACM Symposium on Virtual Reality Software and Technology (ACM)*, 78–81.
- Pelachaud, C. (2009). Studies on Gesture Expressivity for a Virtual Agent. *Speech Commun.* 51, 630–639. doi:10.1016/j.specom.2008.04.009
- Poppe, R., Truong, K. P., Reidsma, D., and Heylen, D. (2010). “Backchannel Strategies for Artificial Listeners,” in *International Conference on Intelligent Virtual Agents*. Berlin, Heidelberg: Springer, 146–158. doi:10.1007/978-3-642-15892-6\_16
- Roseman, I. J. (1991). Appraisal Determinants of Discrete Emotions. *Cogn. Emot.* 5, 161–200. doi:10.1080/02699939108411034
- Rothbaum, B. O., Hodges, L., Smith, S., Lee, J. H., and Price, L. (2000). A Controlled Study of Virtual Reality Exposure Therapy for the Fear of Flying. *J. consulting Clin. Psychol.* 68, 1020–1026. doi:10.1037/0022-006x.68.6.1020
- Shernoff, E. S., Von Schalscha, K., Gabbard, J. L., Delmarre, A., Frazier, S. L., Buche, C., et al. (2020). Evaluating the Usability and Instructional Design Quality of Interactive Virtual Training for Teachers (Ivt-t). *Educ. Tech. Res. Dev.* 68, 3235–3262. doi:10.1007/s11423-020-09819-9
- Shu, Y., Huang, Y.-Z., Chang, S.-H., and Chen, M.-Y. (2019). Do virtual Reality Head-Mounted Displays Make a Difference? a Comparison of Presence and Self-Efficacy between Head-Mounted Displays and Desktop Computer-Facilitated Virtual Environments. *Virtual Reality* 23, 437–446. doi:10.1007/s10055-018-0376-x
- Slater, M., Lott, B., Arnold, M. M., and Sanchez-Vives, M. V. (2009). How We Experience Immersive Virtual Environments: the Concept of Presence and its Measurement. *Anuario de psicología* 40, 193–210.
- Wallach, H. S., Safir, M. P., and Bar-Zvi, M. (2009). Virtual Reality Cognitive Behavior Therapy for Public Speaking Anxiety. *Behav. Modif* 33, 314–338. doi:10.1177/0145445509331926
- Wiederhold, B., and Riva, G. (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, 218. doi:10.3233/978-1-61499-121-2-218

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Glémarec, Lugrin, Bossier, Collins Jackson, Buche and Latoschik. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.