



HAL
open science

Impact of Augmented Engagement Model for Collaborative Avatars on a Collaborative Task in Virtual Reality

Hugo Le Tarnec, Olivier Augereau, Elisabetta Bevacqua, Pierre de Loor

► **To cite this version:**

Hugo Le Tarnec, Olivier Augereau, Elisabetta Bevacqua, Pierre de Loor. Impact of Augmented Engagement Model for Collaborative Avatars on a Collaborative Task in Virtual Reality. AVI 2024: International Conference on Advanced Visual Interfaces 2024, Jun 2024, Arenzano, France. pp.1-9, 10.1145/3656650.3656684 . hal-04620122

HAL Id: hal-04620122

<https://hal.science/hal-04620122v1>

Submitted on 21 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.

Impact of Augmented Engagement Model for Collaborative Avatars on a Collaborative Task in Virtual Reality

Hugo Le Tarnec
b<>com – Institut de recherche technologique &
ENIB, Lab-STICC UMR 6285 CNRS
Brest, France
Hugo.LETARNEC@b-com.com

Elisabetta Bevacqua
ENIB, Lab-STICC UMR 6285 CNRS
Brest, France
elisabetta.bevacqua@enib.fr

Olivier Augereau
ENIB, Lab-STICC UMR 6285 CNRS
Brest, France
olivier.augereau@enib.fr

Pierre De Loor
ENIB, Lab-STICC UMR 6285 CNRS
Brest, France
deloor@enib.fr

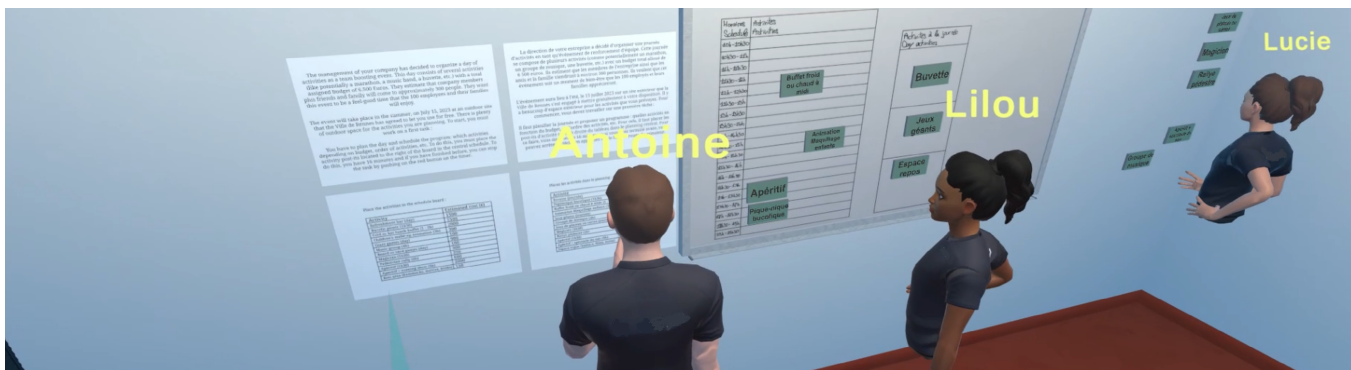


Figure 1: Three participants collaborate remotely on organizing a cultural event in a VR session. The collaborative task involves planning activities for the event within a limited budget. The participants communicate verbally through an audio device and are represented by avatars in the same virtual space.

ABSTRACT

To improve communication and collaboration in virtual reality (VR), we suggest going beyond improving behavioral realism on participant avatars. Leveraging VR capabilities, our approach focuses on transforming and amplifying social interactions. VR technology prevents users from naturally expressing social signals, accurately detecting and transmitting them due to still limited technological capabilities, or even fully perceiving them. We propose an augmented model aimed at enhancing collaboration by displaying non-verbal engagement behaviors on listeners' avatars, even if not executed. In a trio VR session organizing a cultural event, participants were represented by avatars, with or without the augmented model in a within-subject design. Results indicate significantly improved social presence and significant changes in participants' behaviors, particularly their gaze, demonstrating the transformative potential of social augmentations for collaborative tasks in VR.

CCS CONCEPTS

• **Human-centered computing** → **User studies; Empirical studies in collaborative and social computing.**

KEYWORDS

Virtual Reality, Collaboration, Avatar, Non-verbal Behavior, Social Augmentation

ACM Reference Format:

Hugo Le Tarnec, Olivier Augereau, Elisabetta Bevacqua, and Pierre De Loor. 2024. Impact of Augmented Engagement Model for Collaborative Avatars on a Collaborative Task in Virtual Reality. In *International Conference on Advanced Visual Interfaces 2024 (AVI 2024)*, June 3–7, 2024, Arenzano, Genoa, Italy. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3656650.3656684>

1 INTRODUCTION

Considering collaboration as the intersection of coordination, cooperation, and communication [14], it's crucial to acknowledge the significance of each component, with special attention to the challenges posed by communication in the context of VR. Indeed, despite the potential for VR to enhance immersion and presence [37], the communication of humans implies significant signals that are transmitted in real time through non-verbal behaviors that contribute to the sense of social presence [2].

These behaviors involve the full body [41], the head and the face, via head movements [23], facial expressions [21] and gaze [5, 7, 9].

The crucial point is that these signals are subtle and mostly reactive. It is therefore necessary to perceive this non-verbal language in the smallest details for an effective communication because it expresses internal states including the emotional state [11], fatigue [20], stress [15], etc. A frown, for example, is a subtle gesture that can be picked up by an interlocutor to modify his own behavior or explain his thoughts. However, VR technology inhibits users from naturally expressing these social cues, such as nodding, due to the weight of the headset. Then, it is also difficult to detect and transmit them accurately in real time due to the still limited technological capabilities. Finally, due to technological constraints of VR headsets screens (field of view, resolution, etc.), the perception of subtle communication cues from colleagues might be hindered.

Replicating non-verbal behaviors in avatars enhances social presence and collaboration among distant users [31, 39]. Our paper evaluates the hypothesis that triggering non-verbal behaviors at opportune moments, even if not explicitly produced by the avatar user, can further improve collaboration. We introduce “VRinteract”, a semi-autonomous avatar model designed to display non-verbal engagement behaviors on listeners’ avatars, even if not explicitly performed by them. The experimental setup, requiring three users for collaboration, poses technical challenges. Detailed results support the relevance of our proposition concluded with a critical discussion.

2 RELATED WORK

2.1 Social presence

Within collaborative tasks, it is important for collaborators to sense the presence of others, often described as the “feeling of being with another” as defined by Biocca [8]. This notion of “social presence” holds significant relevance in the realm of VR. It is important to note that this feeling is a psychological experience induced by the mediated system and is distinct from immersion, which pertains to the objective qualities brought about by technology within the mediated system, as described by Oh [24]. Due to the fact that social presence often serves as a reliable predictor of positive communication outcomes, both academic researchers and practitioners have expressed considerable interest in investigating factors that contribute to its enhancement. These factors encompass elements like the quality of audio, the resolution of the display, and even individual characteristics such as age and gender [24]. These investigations have revealed how these variables can influence the perception of social presence. One of the most studied factors is the visual representation of the communication partner. Indeed, this factor can lead to better interaction and a change in the behavior of the interlocutors [24]. This factor has demonstrated the potential to facilitate more effective interactions and even influence the behavior of those engaged in communication [24].

Our research is particularly concerned with exploring the concept of social presence and its applicability within our study. This interest stems from the fact that collaboration can be considered as the intersection between communication, coordination, and cooperation [30], and that social presence directly impacts communication.

2.2 Non-verbal behavior

Various aspects of visual representation of the communication partner have been evaluated, including the concept of behavioral appearance, defined as the extent to which the virtual representation behaves as a real person would behave, i.e. replicate the non-verbal behavior of the user. Much work has been dedicated to replicating this behavioral appearance in virtual environments. For example, Wu and his team [41] developed an avatar system for VR collaboration, enabling rendering of users’ body movements, hand gestures and facial expressions using camera tracking technology. Their results showed that their avatars generated more social presence from users and better task performance than less expressive avatars. Moreover, Oh Kruzic et al. [25] showed that participants represented by dynamic avatars with facial expressions and body gestures described their interaction experiences more positively than those without facial expressions and with body gestures during a collaborative task via a 2D screen. Kang and his team [18] have shown that avatars displaying facial expressions elicit a stronger feeling of social presence compared to static avatars.

2.3 Transformed Social Interaction

Numerous studies focus on replicating non-verbal communication in virtual and VR environments. However, VR’s potential allows for artificial modification of social interactions beyond face-to-face experiences. Hollan and Stornetta [17] argue that non-imitative methods may offer better solutions than merely mimicking face-to-face interactions.

Collaborative Virtual Environments (CVEs) uniquely enable the filtering, amplification, and transformation of non-verbal behaviors during social interactions. Filtering separates behaviors that facilitate from behaviors that degrade social interactions, and CVEs enable a deliberate disconnection between displayed and actual behaviors. These dissociation techniques can either enhance or diminish aspects of non-verbal communication in CVEs, offering participants the opportunity to benefit from enhanced non-verbal cues or experience reduced non-verbal communication. Determining which cues constitute a more positive or negative interaction in this context is challenging, as non-verbal behaviors are typically assessed within a broader, unconscious, and subtle context. Consequently, there is a gap in understanding how social interactions in VR environments may evolve as we adapt to and fully exploit the potential of these immersive environments.

Bailenson and his team [4] introduced the notion of “Transformed Social Interaction” (TSI), involving generating non-verbal behaviors in virtual environments that differ from behaviors exhibited in reality. Bailenson classifies TSIs into three categories: “Transforming Sensory Capabilities”, “Transforming the Situation” and “Transforming Self Representations”.

The first category, “Transforming Sensory Capabilities”, aims to supplement limited perceptual capacities by providing real-time summary information about the attentions and movements of oneself and others. Non verbal gestures are often correlates of specific mental states [1, 12, 43]. For example, integrating a display indicating the time of mutual gaze with different participants can help adapt behavior accordingly. Bailenson and his team [6] demonstrated that teachers, given visual warnings about students not

receiving adequate gaze, exhibited greater ability to equitably allocate attention compared to teachers who did not receive such warnings.

The second category, “Transforming the Situation”, involves changing the spatial or temporal structure of a conversation. For example, the participants in the discussion are placed at equal distance to improve the discussion. Bailenson and his team [6] conducted a study in which they noted that students situated at the periphery of a classroom tended to receive less visual attention from the teacher compared to those positioned in the central area of the room. Consequently, they evaluated the impact of allowing two students to occupy a single favored seat while one of them perceived the other as being located elsewhere. The findings of this study indicated that there was an increase in learning outcomes when this modified proximity arrangement was implemented.

The third category, “Transforming Self Representations”, allows avatars in CVEs to be different from users, whether in photographic appearance [42] or behavior. The modification of the behavioral self-representation consists of having avatar non-verbal behavior different from the user real behavior. Hybrid systems combining user and artificial behaviors have emerged and can enhance interaction by mixing a real user model with a social computing model. Roth et al. [32] developed a hybrid gaze model combining a natural user-derived model with an automatic social gaze model. In addition to models blending user behavior reproduction with artificial behaviors, there are those merging user behavior with their conversation partner. Roth and his team [33] studied a hybrid prototype modifying upper body movement based on non-verbal mimicry between users, demonstrating its positive impact on social interactions. Thus, to improve communication and collaboration in VR, TSI could go beyond a simple duplication of user behavior. Conversational engagement, a key concept for collaboration [35], could be increased positively. Poggi and her team [28] define conversational engagement as the value that a participant attributes to the goal of being together with others and continuing the interaction. However, in VR, detecting, transmitting, and perceiving these non-verbal engagement signals can be challenging.

2.4 Conversational engagement

To perform collaborative work in VR, it is necessary that there is a certain level of engagement between collaborators. Recent studies on virtual agents and communication robots have revealed that conversational engagement is fundamental and indispensable in the communication between human users and humanoid interfaces [27]. Indeed, as Sidner and colleagues argue [34], evidence of the importance of engagement becomes verifiable in situations where non-verbal engagement behaviors are in conflict with verbal behaviors and lead to an impossibility of maintaining communication. The issue is that the conversational non-verbal engagement behaviors made by the listeners and rendered identically on their avatar are not very perceptible by the speaker in VR. In fact, the precise detection of micro-expressions of engagement made by listeners in VR is difficult from a technological point of view and still a current research issue [16, 26, 38]. Similarly, the weights of VR headsets lead to reduced mobility, and therefore, less head-nods. Moreover, the proxemic behaviors and the orientations of the body

in the physical world are partly explained by the need for perception of the speech of the interlocutors. To perceive the speech of an interlocutor, we tend to turn towards them and reduce our interpersonal distance. Thus, in the case where the sound of speech is not spatialized, these behaviors can be modified.

In consequence, we aim to develop a hybrid model based on TSI, displaying artificial non-verbal engagement behaviors on listeners’ avatars, even without actually being executed by themselves. In this work, we examine the influence of an augmented engagement model for collaborative tasks in VR. We hypothesize that our *augmented engagement model improves collaboration in VR*. As mentioned previously, related work has explored the effect of a hybrid model [32] to actively mediate communication in VR. Our work is similar to this, but it focuses on generating varied engagement behaviors for collaboration. Moreover, the non-verbal behaviors of the listener avatar are not only a blend of the actual non-verbal behavior of the listener and non-verbal behavior that an agent could exhibit. Instead, they derive additionally from the gaze behaviors of the speaker. Finally, we propose a technically challenging experimental setup with a trio collaboration task.

3 VRINTERACT

In this section, we will explore the operational principles of our model and how it adapts to the intricacies of dynamic collaborations in VR.

When individuals engage in collaborative tasks in VR, they can dynamically take on the roles of speaker, indicating that they are actively speaking, or listeners, indicating that they are actively listening to someone else speaking. It’s worth noting that in scenarios where multiple individuals are simultaneously speaking, some participants may hold both roles concurrently. Conversely, in scenarios where no one is actively speaking, users do not adopt any particular role.

In essence, our model is designed to generate artificial non-verbal behaviors for the avatars of listeners seen by a speaker, thereby enhancing their collaborative experience. However, for those who are not designated as listeners, their avatars remain unaltered and faithfully reproduce the real behaviors of the users.

Specifically, the model offers the ability to dynamically switch between three distinct states for listener avatar seen by a speaker: “mirror”, “shared gaze object”, and “mutual gaze”, as depicted in Figure 2. This model is based on the “gaze-following” model developed by Frädrieh and his research team, as detailed in their work [13]. The differences are that our model generates non-verbal behaviors for an avatar of a real listener, not just focusing on gaze but a range of non-verbal cues. Lastly, our model is designed to adapt to the specific dynamics of collaborative tasks, where collaborators can dynamically switch between the roles of speaker and listener. By default, all listener avatars observed by a speaker are set to the “mirror” state, replicating as much as possible the users behavior.

“Mirror” State (by default): This state replicates the listener’s non-verbal behaviors as closely as possible without adding any artificial non-verbal behaviors. It is controlled in inverse kinematics from the VR headset for the position of the head and two controllers for the position of the hands. In addition, facial expressions are

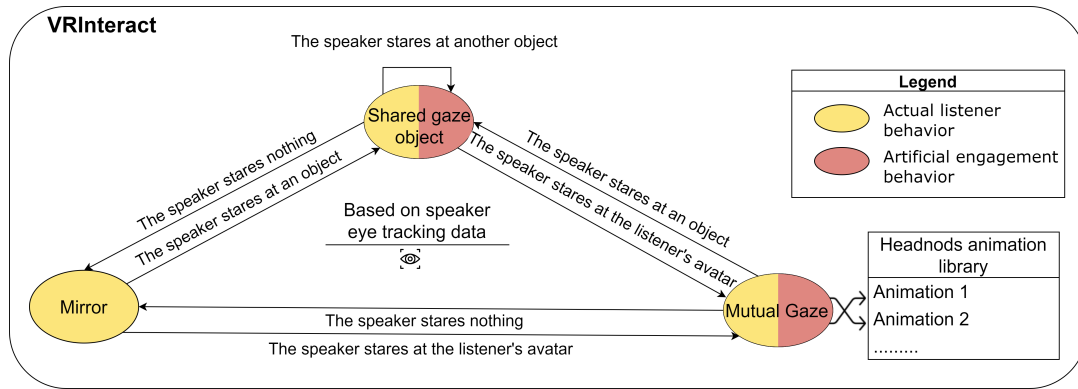


Figure 2: VRInteract model for generating the non-verbal behavior of listener avatar seen by a speaker in a collaborative task in VR. It consists of 3 states: “mirror”, “shared gaze object” and “mutual gaze”. The default state is “mirror” where the non-verbal behavior of the listener avatar is reproduced identically. The avatar can look at the speaker’s object of focus (“Shared Gaze Object”) or look at the speaker when the speaker is looking at the listener’s avatar (“Mutual Gaze”). This approach aims to maintain authenticity of interaction and engagement in collaborative VR tasks.

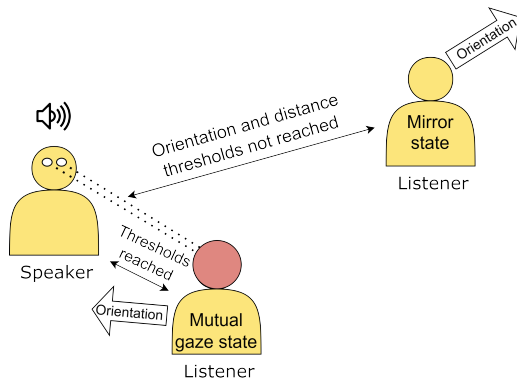


Figure 3: In this example during the collaborative task, one participant is the speaker, while the other two are listeners. The speaker looks at the avatar of listener 1, and as orientation and distance conditions are met, listener 1’s avatar enters the “Mutual gaze” state (visible only to the speaker). Listener 1 starts looking at the speaker, accompanied by head movements and engaging facial expressions. The other listener’s avatar remains in a mirror state since the conditions for orientation, distance, and being looked at by the speaker are not met. The model enables real-time adjustments to avatar non-verbal behaviors to accommodate dynamic collaboration in VR, including role transitions like speaker to listener.

copied identically based on those of the listener. You can find a comprehensive description of its implementation in Section 4.3.

“Shared Gaze Object” State: In this state, when the speaker stares at an object, the listener’s avatar synchronizes its gaze with the speaker’s selected object, even if the listener is not actually focusing on it (see Figure 4). Artificial engagement behaviors to maintain interaction focus exclusively on head and face movements.



Figure 4: On the left, the generated behavior of shared gaze on an object : the listener’s avatar begins to look at the object looked at by the speaker, even if the listener is not actually looking at it. On the right, the avatar is in mirror state and no behavior is increased.

Indeed, in addition to changing the direction of the head, artificial facial expressions of engagement are displayed on the avatar. User proxemic behaviors and gestures are retained to ensure credibility, especially during interactions involving virtual objects in collaborative VR tasks. Actually, keeping the real position of the body and hands allows consistency during manipulation phases. Additionally, to reach this state, distance and angle thresholds, synonymous with conversational attention cues, are employed to gauge the listener’s level of the attention towards the speaker. Activation of the “Shared Gaze Object” state occurs only when the listener is sufficiently close and correctly oriented towards the speaker. These thresholds are established through empirical investigation during the pre-test phase and may exhibit variability contingent upon factors such as the specific task, the environment, or the particular VR hardware employed, among other potential variables. The specification of the threshold values for this study is explicated within Section 4.3.

“Mutual Gaze” State: In this state, when the speaker is looking at the listener avatar, the listener’s avatar maintains eye contact with the speaker. This state shows the significance of the face as a

crucial channel in non-verbal communication and its role in regulating conversation flow [10]. Concurrently, the avatar incorporates engaging head movements and facial expressions (see Figure 5). These head movements, such as nods and jerks, serve to provide backchannels to maintain interaction continuity, and feedback for perception and understanding of the message [3, 36]. Additionally, facial expressions, including smiles, can encourage speakers to continue the conversation [29]. Distance and angle thresholds are also employed to gauge the listener’s level of attention towards the speaker. Likewise, the activation of the “Mutual Gaze” state occurs only when the listener is sufficiently close and correctly oriented towards the speaker.

In summary, the default state is “mirror”. The model adjusts the listener’s avatar behavior seen by the speaker based on the conversational attention cues of the listener, which encompass factors such as distance and orientation relative to the speaker. Depending on these cues, the avatar may stare at the speaker’s object of focus (“Shared Gaze Object”) or look at the speaker when the speaker gazes at the listener’s avatar (“Mutual Gaze”). When the conditions for the above states are not met, the avatar returns to its default “mirror” state. A scenario is presented in Figure 3 to better understand the use of the model. This approach aims to maintain interaction authenticity and engagement in collaborative VR tasks.

4 METHODS AND MATERIALS

4.1 Design and Hypotheses

This study was conducted according to a within-subjects design comparing the conditions “Mirror Avatars”(“MA”) and “Augmented Avatars”(“AA”). “MA” are avatars that closely reproduce the user’s actual behavior on their avatar, i.e. remaining exclusively in the “mirror” state. “AA” are avatars using the “VRInteract” model. Participants collaborated in groups of three during a VR session to collectively organize a cultural event. The participants were located into three different rooms. The participants were blind to the real goal of the experiment.

Our hypotheses are:



Figure 5: On the left, the generated mutual gaze behavior: the listener’s avatar starts looking at the speaker’s avatar, accompanied by head nods and facial expressions of engagement, even if the listener is not actually looking at it. On the right, the avatar is in mirror state and no behavior is increased.

H1: *Participants using “AA” experience more social presence than those using “MA”.*

H2: *Participants using “AA” experience more teamwork satisfaction than those using “MA”.*

H3: *“AA” induces more conversational attention between participants compared to “MA”.*

4.2 Task

In all the experiments mentioned above [18, 25, 41], the tasks were chosen to promote face-to-face interactions without involving the manipulation of virtual objects. Indeed, this type of task is a factor of social presence [24]. But, in general, when users engage in collaborative work in VR, they will often need to interact with the virtual environment and its various objects that compose it. Consequently, we have chosen as a collaborative task, to collectively organize a cultural event during a VR session.

The participants were grouped into trio to do the task divided into two subtasks:

- Day planning task (see Figure 1): the task aims to organize a day broken down into several activities (such as a music group, a refreshment bar, etc.) with a total allocated budget of 6,500 euros. The task is to choose the desired activities with costs defined in a planning table respecting the total budget, the planned date (in summer), the location (outdoors). Participants should use sticky notes corresponding to the different activities located to the right of the board and place them on the whiteboard. When the participants redo the task for the second condition, the budget (4000 euros), the date (in winter), and the location (indoors) have changed. In addition, they must organize an environmentally friendly event.
- Slogan creation task: the purpose of the task is to find titles for the event and associated slogans, and rank them by preference. Participants should use voice sticky notes located to the right of the board and place them on the whiteboard.

4.3 Implementation

The virtual environment is created using the Nengagement 1.0 framework, which was created in Unity 3D by B-com¹. This framework enables multiple users to experience a VR shared environment at the same time. For the environment, we had to model the room in 3D with the whiteboard. In addition, various features have been developed such as voice sticky notes, using speech-to-text technology, allowing you to write on the notes using voice commands.

For the implementation of “MA” and “AA”, the avatars are controlled in inverse kinematics (IK) using the HTC Vive Pro VR headset for head position and two HTC controllers for hand position. Ready Player Me avatars² were chosen because they feature blend-shapes for character animations and facial expressions.

For “MA”, the aim was to faithfully reproduce the user’s non-verbal behaviors in the avatar. Facial expressions, based on 41 key points collected by an HTC Facial Tracker sensor and an eye tracker in the HTC Vive Pro Eye, were displayed in real time. Each key point

¹B-com website <https://b-com.com/en>

²Metaverse Complete Online 3D Avatar Creator – Ready Player Me. <https://readyplayer.me/>.

was associated with a blendshape value to animate the avatar’s face. Gaze direction was replicated from the user using the eye tracker. A calibration phase was required for rendering facial expressions and gaze direction.

For “AA”, based on our model “VRInteract”, avatar state is managed in the animator controller on Unity. When avatars are in the “mirror” state, they behave like “MA”. When avatars are in “shared gaze object” state, the avatar turns its head to look at the object looked at by the speaker, detected with its eye tracker. When avatars are in “mutual gaze” mode, the avatar turns toward the listener accompanied by head movements and facial expressions of engagement. Concretely, our model randomly activates, from a library of predefined animations, an engagement head movement animation combined with an engagement facial animation.

The determination of orientation and distance threshold parameters in the “VRInteract” model for initiating artificial behaviors was conducted through empirical analysis in the pre-test phase. We opted for a specific angle value of 55°, which aligns with the angular extent of the HTC Vive Pro headset’s field of view (110° divided by 2, resulting in 55°). When the calculated angle fell below 55°, it signified that the speaker was situated within the field of view of the listener.

4.4 Setup and apparatus

For each participant, an HTC Vive Pro Eye tracker headset with an HTC Facial Tracker sensor, with two HTC Vive Pro joystick controllers, two HTC VIVE Pro 2.0 bases and one computer were used. In addition, for the server, one computer was used. Thus, in total, three HTC Vive Pro headsets, three HTC Facial Trackers, six controllers, six HTC Station bases and four computers were used.

4.5 Procedure

Participants in the study provided informed consent and completed demographic and technology questionnaires. After receiving scripted oral instructions by the experimenter, each participant filled out the Simulator Sickness Questionnaire (SSQ) [19] and was then equipped with a VR headset and two joysticks.

A three-minute practice session allowed participants to familiarize themselves with controllers. In the first condition, participants entered a virtual testing environment, starting data recording when they pressed the timer start button. The trio had 16 minutes for a day planning subtask, followed by 8 minutes for a slogan creation subtask. Data logging started and stopped with the timer. Participants could signal the experimenter for technical issues.

Post-task, participants removed VR headsets to answer the SSQ, social presence [22], and teamwork satisfaction questionnaire [40]. The second condition involved re-equipping and data recording, with tasks modified in theme, date, and budget. Post-task, participants filled out the same questionnaires. The entire experiment took about an hour and occurred at the European Center for Virtual Reality³.

4.6 Measures

4.6.1 Subjective Measures. For *H1*, social presence was evaluated using a five-question questionnaire (5-point Likert scale) based on

³European Center for Virtual Reality <https://cerv.enib.fr>

Makransky et al.’s multimodal presence questionnaire [22]. The questionnaire was administered after each condition.

For *H2*, teamwork satisfaction was measured using a ten-question questionnaire [40] (5-point Likert scale) to assess the impact of the augmented engagement model. The questionnaire was administered after each condition.

4.6.2 Objective Measures. To investigate *H3*, we collected three types of behavioral data related to conversational attention for each participant: gaze behavior, the participant’s orientations relative to other avatars, and the participant’s distances relative to other avatars. The analysis specifically focused on periods when the participant was speaking, as our model is only active during these speaking instances.

For gaze behaviors, we used eye tracking to determine where the participant was actually looking. We quantified “avatar looking time” as the percentage of time spent looking at the faces of other avatars while speaking, divided by their total speaking time.

Participants’ orientations and distances were directly collected from the virtual environment at a frequency of 5 Hz. This lower frequency was sufficient for these behavioral measures. Two measures were calculated: “minimum orientation”, representing the average minimum orientation of the speaker towards each participant every 0.2 seconds (5 Hz) while speaking, divided by their total speaking time and the “minimum interpersonal distance”, indicating the average minimum distance between the speaker and each participant every 0.2 seconds (5 Hz) while they speak, divided by their total speaking time.

4.6.3 Control Measures. To assess whether other factors influence the subjective and objective measures, we administered pre-study questionnaires with a demographic questionnaire, a technology experience and familiarity questionnaire. In addition, we administered the Simulator Sickness Questionnaire [19] before and immediately after each donning of the VR headset.

4.7 Participants

Participants (N=30) were recruited through university and company mailing lists, forming 10 groups for the experiment. All analyses were conducted using the data from these 30 participants, except for the behavioral data analysis, which excluded one trio due to a technical issue. So, for hypothesis *H3*, data from 27 participants were used. On average, participants spend around 10 hours per day in front of a screen. The sample consisted of a distribution of 16 women and 14 men, with an average age of 30.20 years.

5 RESULTS

5.1 Descriptive data

On average, the duration of the day planning task lasted 11 minutes and 17 seconds (16 min allocated) and the duration of the slogan task lasted 7 minutes and 33 seconds (8 min allocated). Participants spoke an average of 68% of the time on the different tasks ($M = 0.68$, $SD = 0.15$, “M” represents the mean and “SD” represents the standard deviation).

During a participant’s speaking time, artificial behaviors are triggered approximately 64% of the time on the avatar of these listeners (i.e. 44% of the total task time), with a distribution of 10%

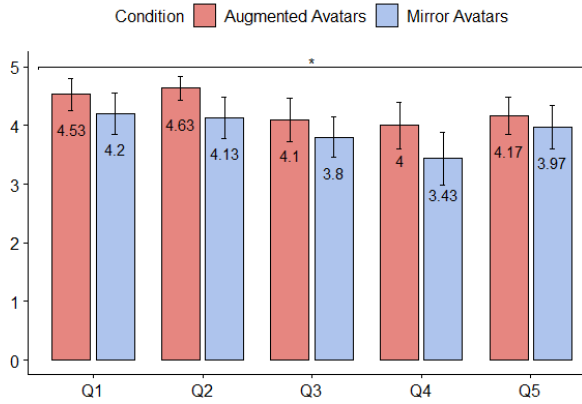


Figure 6: Result of the social presence questionnaire with the averages of the 5 questions per condition. * indicates $p < .05$ for the Wilcoxon-Mann-Whitney test.

of mutual gaze and 54% shared gaze on an object (respectively 7% and 37% of the total time spent on the task).

5.2 Subjective results

To test hypothesis $H1$ regarding the impact of “Augmented Avatars” compared to “Mirror Avatars” on social presence, we employed a Wilcoxon signed-rank test on the averages of the 5 questions from the social presence questionnaire (Section 4.6.1). Results revealed a significant difference in perceived social presence between the conditions “Mirror avatar” ($M = 3.91$, $SD = 0.75$, “M” represents the mean and “SD” represents the standard deviation) and “Augmented Avatar” ($M = 4.29$, $SD = 0.58$, $p = .049$) (see Figure 6).

To test hypothesis $H2$ and see the impact of “Augmented Avatars” compared to “Mirror Avatars” on teamwork satisfaction, we conducted a Wilcoxon signed-rank test on the averages of the 10 questions of the teamwork satisfaction questionnaire presented in the Section 4.6.1. The results showed a non significant difference in perceived teamwork satisfaction between the conditions “Mirror avatar” ($M = 3.33$, $SD = 0.41$) and “Augmented Avatar” ($M = 3.54$, $SD = 0.33$, $p = 0.052$)

5.3 Objective results

We chose to combine the objective data of the subtasks (day planning task and slogan creation task) because there was no significant difference in the two.

To test hypothesis $H3$ and see the impact of augmented avatar compared to mirror avatar on conversational attention, and specifically on “gaze time on avatar”, we conducted paired samples t-tests, in compliance with the test application conditions. The results showed a significant difference in the conditions “Mirror avatar” ($M = 0.21$, $SD = 0.10$) and “Augmented Avatar” ($M = 0.27$, $SD = 0.12$, $t(50.74) = -2.05$, $p = 0.046$). We conducted paired samples t-tests, in compliance with the test application conditions, to see the impact on conversational attention through proxemic behaviors, specifically on “minimum interpersonal distance”. Results showed no significant difference between the conditions “Mirror avatar”

($M = 1.03$, $SD = 0.231$) and “Augmented Avatar” ($M = 1.06$, $SD = 0.204$, $t(51.23) = -0.4$, $p = 0.69$). We conducted paired samples t-tests, in compliance with the test application conditions, to see the impact on conversational attention through orientation behaviors, specifically on “minimum orientation”. Results showed no significant difference between the conditions “Mirror avatar” ($M = 50.39$, $SD = 10.17$) and “Augmented Avatar” ($M = 50.23$, $SD = 12.94$, $t(49.26) = 0.05$, $p = 0.96$).

5.4 Control results

To ensure no bias in both conditions, we conducted paired samples t-tests for the total simulator sickness score (TS), comprising nausea (N), oculomotor disturbance (O), and disorientation (D) sub-scales. No significant difference was found between the two conditions and before/after VR headset use. Participants in both conditions experienced no or mild symptoms of cybersickness.

6 DISCUSSION AND FUTURE WORK

Below we present a discussion of the results, some limitations of the current study, potential future work, and broader reflections. We investigated how an augmented engagement model could impact collaboration in VR.

Regarding hypothesis $H1$, we found a significantly higher perception of social presence in the “Augmented Avatars” condition compared to the “Mirror Avatars” condition. This implies that participants observing artificial behaviors on the listeners’ avatars, both mutual gaze and both shared gaze at objects, will experience more of the feeling of being with another.

Regarding $H2$, we did not observe a significantly higher perception of teamwork satisfaction in the “Augmented Avatars” condition compared to the “Mirror Avatars”. This suggests that participants using the augmented avatars may not necessarily experience greater teamwork satisfaction when compared to participants using the mirror avatars. Nevertheless, for item 4: “I have benefited from interacting with my teammates”, we found a significant difference between the two conditions ($p = 0.049$). Likewise, for item 5: “I have benefited from my teammates’ feedback”, we found a significant difference between the two conditions ($p = 0.034$). This shows that artificial engagement behaviors generated visual feedback making it possible to maintain interaction and have benefits for the participants (see Figure 7).

In relation to $H3$, a significant distinction was observed in the percentage of time spent looking at other avatars’ faces between the “Augmented Avatars” and “Mirror Avatars” conditions. Participants in the “Augmented Avatars” condition exhibited a higher tendency to gaze at other avatars’ faces compared to the “Mirror Avatars” condition. This indicates that augmented avatars induce a shift in participant behavior, specifically in gaze behavior, indicative of conversational attention. However, no significant differences were identified in orientation and proxemic behavior, both reflective of conversational attention, between the two conditions.

Hence, $H1$ is validated, and $H3$ is partially validated, especially concerning gaze behavior, indicative of conversational attention. Participants with “augmented avatar” reported higher social presence and exhibited more extended gaze at other avatars’ faces compared to those with “mirror avatars”. $H2$ is also partially confirmed, with significance observed in two out of ten items.

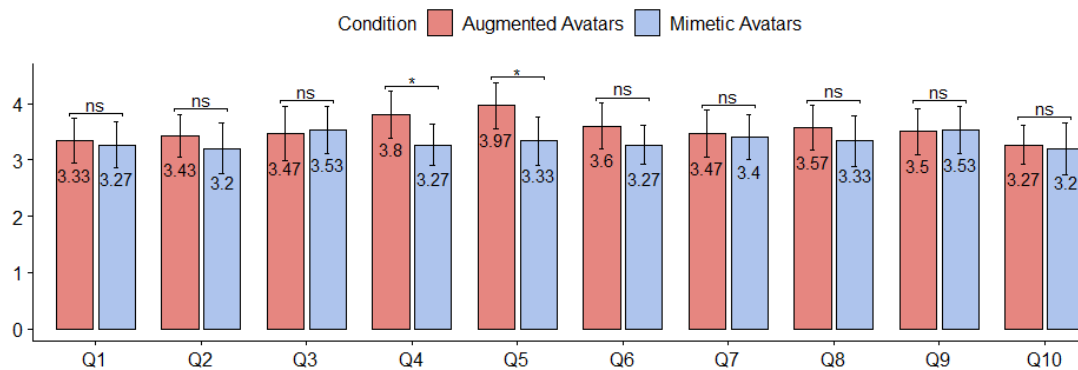


Figure 7: Result of the teamwork satisfaction questionnaire. For the item 4: “I have benefited from interacting with my teammates”, we found a significant difference between the two conditions ($p = 0.049$). Likewise, for item 5: “I have benefited from my teammates’ feedback”, we found a significant difference between the two conditions ($p = 0.034$). * indicates $p < .05$ and “ns” indicates $p > .05$ for the Wilcoxon-Mann-Whitney test.

With our main research goal, we demonstrate the possibility of enhancing collaboration in VR experiences by transforming and augmenting social interactions. These findings underscore the broad applicability of social enhancements in various collaborative VR scenarios. However, one group exhibited primarily negative interaction outcomes, characterized by frequent disagreement in decision-making and limited participant engagement. Despite this, artificial behaviors were consistently triggered, occasionally conflicting with the listener’s cognitive engagement. Furthermore, it’s essential to assess the model’s performance in scenarios involving more than three collaborators to determine its generalizability in such contexts.

In our future work, we aim to refine the model by developing a more accurate measure of listeners’ engagement towards speakers, allowing for the activation of artificial behaviors. Currently, these behaviors are initiated based on predefined conditions, such as listener proximity and orientation. While these conditions are determined empirically, they may not fully capture the complexity of human interaction. Our aim is to use a wider range of input data, including facial expressions, head movements, orientation, distance, and gaze, to understand listeners’ conversational engagement better. We plan to communicate this information in various ways. Currently, the artificial engagement behaviors we generate imitate real participant actions. However, there’s potential to improve user interaction and collaborative dynamics by presenting engagement information through augmented visuals in VR or using alternative sensory channels like auditory or tactile feedback.

7 CONCLUSION

In this paper, we presented the design, implementation and an evaluation of a concept for the improvement of collaborative work in VR. We propose an augmented model for displaying non-verbal engagement behaviors on listener’s avatars, even if they are not actually executed. The model, called “VRInteract” is hybrid, mixing the reproduction of the user’s non-verbal behaviors with the addition of artificial behaviors. To evaluate this model, we developed an environment where participants collaborate as a trio to organize a cultural event during a VR session. We hypothesized that the use

of avatars using “VRInteract” improves collaborative work in terms of social presence, teamwork satisfaction and leads to changes in conversational attention behaviors compared to the use of mirror avatars which reproduces the non-verbal behavior of the user as closely as possible.

The results of the social presence questionnaire and the results of gaze behavior, related to conversational attention showed a notable and statistically significant difference between the conditions “Mirror Avatars” and “Augmented Avatars” but non significant difference between the two conditions for the teamwork satisfaction questionnaire. Our results extend the results of previous work [4] and suggest that applied augmentations can have a significant impact on users’ social presence and behavior. We believe that our approach constitutes a first step to explore the potentials of VR as a means of actively mediating collaboration via augmenting the concept of conversational engagement.

In future research, our goal is to enhance the model by integrating a more refined metric for assessing the degree of conversational engagement displayed by listeners towards speakers, with the intention of more precisely activating artificial engagement behaviors.

ACKNOWLEDGMENTS

This research is conducted under the eXtendR project, funded by the French National Agency for Research (ANR) and hosted at the b<>com research institute. We appreciate b<>com’s valuable support and acknowledge the CONTINUUM project for supplying the necessary materials and equipment for our experiments.

REFERENCES

- [1] Evgeniy Abdulin and Oleg Komogortsev. 2015. User eye fatigue detection via eye movement behavior. In *Proceedings of the 33rd annual ACM conference extended abstracts on human factors in computing systems*. 1265–1270.
- [2] Katrin Allmendinger. 2010. Social presence in synchronous virtual learning situations: The role of nonverbal signals displayed by avatars. *Educational Psychology Review* 22, 1 (2010), 41–56.
- [3] Jens Allwood and Loredana Cerrato. 2003. A study of gestural feedback expressions. In *First nordic symposium on multimodal communication*. Copenhagen, 7–22.
- [4] Jeremy N Bailenson, Andrew C Beall, Jack Loomis, Jim Blascovich, and Matthew Turk. 2004. Transformed social interaction: Decoupling representation from

- behavior and form in collaborative virtual environments. *Presence: Teleoperators & Virtual Environments* 13, 4 (2004), 428–441.
- [5] Jeremy N Bailenson, Jim Blascovich, Andrew C Beall, and Jack M Loomis. 2001. Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. *Presence: Teleoperators & Virtual Environments* 10, 6 (2001), 583–598.
 - [6] Jeremy N Bailenson, Nick Yee, Jim Blascovich, Andrew C Beall, Nicole Lundblad, and Michael Jin. 2008. The use of immersive virtual reality in the learning sciences: Digital transformations of teachers, students, and social context. *The Journal of the Learning Sciences* 17, 1 (2008), 102–141.
 - [7] Gary Bente, Felix Eschenburg, and Lisa Aelker. 2007. Effects of simulated gaze on social presence, person perception and personality attribution in avatar-mediated communication. In *Presence 2007: Proceedings of the 10th Annual International Workshop on Presence, October 25-27, 2007, Barcelona, Spain*. Citeseer, 207–14.
 - [8] Frank Biocca, Chad Harms, and Judee K Burgoon. 2003. Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence: Teleoperators & virtual environments* 12, 5 (2003), 456–480.
 - [9] Sara Dalzel-Job. 2015. Social interaction in virtual environments: the relationship between mutual gaze, task performance and social presence. (2015).
 - [10] Starkey Duncan. 1974. On the structure of speaker–auditor interaction during speaking turns. *Language in society* 3, 2 (1974), 161–180.
 - [11] Paul Ekman. 1993. Facial expression and emotion. *American psychologist* 48, 4 (1993), 384.
 - [12] Paul Ekman and Wallace V Friesen. 1978. Facial action coding system. *Environmental Psychology & Nonverbal Behavior* (1978).
 - [13] Laura Frädrieh, Fabrizio Nunnari, Maria Staudte, and Alexis Heloir. 2018. (Simulated) listener gaze in real-time spoken interaction. *Computer Animation and Virtual Worlds* 29, 3-4 (2018), e1831.
 - [14] Hugo Fuks, Alberto Raposo, Marco A Gerosa, Mariano Pimental, and Carlos JP Lucena. 2008. The 3c collaboration model. In *Encyclopedia of E-collaboration*. IGI Global, 637–644.
 - [15] Dimitris Giakoumis, Anastasios Drosou, Pietro Cipresso, Dimitrios Tzovaras, George Hassapis, Andrea Gaggioli, and Giuseppe Riva. 2012. Using activity-related behavioural features towards more effective automatic stress detection. (2012).
 - [16] Steven Hickson, Nick Dufour, Avneesh Sud, Vivek Kwatra, and Irfan Essa. 2019. Eyemotion: Classifying facial expressions in VR using eye-tracking cameras. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 1626–1635.
 - [17] Jim Hollan and Scott Stornetta. 1992. Beyond being there. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 119–125.
 - [18] Sin-Hwa Kang, James H Watt, and Sasi Kanth Ala. 2008. Communicators’ perceptions of social presence as a function of avatar realism in small display mobile communication devices. In *Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008)*. IEEE, 147–147.
 - [19] Robert S Kennedy, Norman E Lane, Kevin S Berbaum, and Michael G Lienthal. 1993. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology* 3, 3 (1993), 203–220.
 - [20] Jarek Krajewski, David Sommer, Udo Trutschel, Dave Edwards, and Martin Golz. 2009. Steering wheel behavior based estimation of fatigue. In *Driving assessment conference*, Vol. 5. University of Iowa.
 - [21] Eva Krumhuber, Antony SR Manstead, Darren Cosker, Dave Marshall, Paul L Rosin, and Arvid Kappas. 2007. Facial dynamics as indicators of trustworthiness and cooperative behavior. *Emotion* 7, 4 (2007), 730.
 - [22] Guido Makransky, Lau Lilleholt, and Anders Aaby. 2017. Development and validation of the Multimodal Presence Scale for virtual reality environments: A confirmatory factor analysis and item response theory approach. *Computers in Human Behavior* 72 (2017), 276–285.
 - [23] Catharine Oertel, José Lopes, Yu Yu, Kenneth A Funes Mora, Joakim Gustafson, Alan W Black, and Jean-Marc Odobez. 2016. Towards building an attentive artificial listener: On the perception of attentiveness in audio-visual feedback tokens. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*. 21–28.
 - [24] Catherine S Oh, Jeremy N Bailenson, and Gregory F Welch. 2018. A systematic review of social presence: Definition, antecedents, and implications. *Frontiers in Robotics and AI* (2018), 114.
 - [25] Catherine Oh Kruzic, David Kruzic, Fernanda Herrera, and Jeremy Bailenson. 2020. Facial expressions contribute more than body movements to conversational outcomes in avatar-mediated virtual environments. *Scientific reports* 10, 1 (2020), 20626.
 - [26] Kyle Olszewski, Joseph J Lim, Shunsuke Saito, and Hao Li. 2016. High-fidelity facial and speech animation for VR HMDs. *ACM Transactions on Graphics (TOG)* 35, 6 (2016), 1–14.
 - [27] Christopher Peters. 2005. Direction of attention perception for conversation initiation in virtual environments. In *Intelligent Virtual Agents: 5th International Working Conference, IVA 2005, Kos, Greece, September 12-14, 2005. Proceedings 5*. Springer, 215–228.
 - [28] Isabella Poggi. 2007. Mind, hands, face and body. *A goal and belief view of multimodal communication*. Weidler, Berlin (2007).
 - [29] Isabella Poggi and Catherine Pelachaud. 2000. Performative facial expressions in animated faces. *Embodied conversational agents* (2000), 155–189.
 - [30] Sam Redfern and Niall Galway. 2002. Collaborative virtual environments to support communication and community in internet-based distance education. *Journal of Information Technology Education. Research* 1 (2002), 201.
 - [31] Shane L Rogers, Rebecca Broadbent, Jemma Brown, Alan Fraser, and Craig P Speelman. 2022. Realistic motion avatars are the future for social interaction in virtual reality. *Frontiers in Virtual Reality* 2 (2022), 750729.
 - [32] Daniel Roth, Peter Kullmann, Gary Bente, Dominik Gall, and Marc Erich Latoschik. 2018. Effects of hybrid and synthetic social gaze in avatar-mediated interactions. In *2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 103–108.
 - [33] Daniel Roth, David Mal, Christian Felix Purps, Peter Kullmann, and Marc Erich Latoschik. 2018. Injecting nonverbal mimicry with hybrid avatar-agent technologies: A naïve approach. In *Proceedings of the 2018 ACM Symposium on Spatial User Interaction*. 69–73.
 - [34] Candace L Sidner, Christopher Lee, Cory D Kidd, Neal Lesh, and Charles Rich. 2005. Explorations in engagement for humans and robots. *Artificial Intelligence* 166, 1-2 (2005), 140–164.
 - [35] Candace L Sidner, Christopher Lee, and Neal Lesh. 2003. Engagement when looking: behaviors for robots when collaborating with people. In *Diabrock: Proceedings of the 7th workshop on the Semantic and Pragmatics of Dialogue*. Citeseer, 123–130.
 - [36] M Skhiri and L Cerrato. 2003. Analysis and measurement of communicative gestures in human dialogues.
 - [37] Mel Slater. 2018. Immersion and the illusion of presence in virtual reality. *British journal of psychology* 109, 3 (2018), 431–433.
 - [38] Katsuhiko Suzuki, Fumihiko Nakamura, Jiu Otsuka, Katsutoshi Masai, Yuta Itoh, Yuta Sugiura, and Maki Sugimoto. 2017. Recognition and mapping of facial expressions to avatar by embedded photo reflective sensors in head mounted display. In *2017 IEEE Virtual Reality (VR)*. IEEE, 177–185.
 - [39] Hugo Le Tarnec, Elisabetta Bevacqua, Olivier Augereau, and Pierre De Loor. 2023. Effect of Avatar Facial Expressiveness on Team Collaboration in Virtual Reality. In *Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents*. 1–8.
 - [40] Hungwei Tseng, C Wang, H Ku, and Ling Sun. 2009. Key factors in online collaboration and their relationship to teamwork satisfaction. *Quarterly Review of Distance Education* 10, 2 (2009), 195–206.
 - [41] Yuanjie Wu, Yu Wang, Sungchul Jung, Simon Hoermann, and Robert W Lindeman. 2021. Using a fully expressive avatar to collaborate in virtual reality: Evaluation of task performance, presence, and attraction. *Frontiers in Virtual Reality* 2 (2021), 641296.
 - [42] Nick Yee and Jeremy Bailenson. 2007. The Proteus effect: The effect of transformed self-representation on behavior. *Human communication research* 33, 3 (2007), 271–290.
 - [43] Robert B Zajonc, Sheila T Murphy, and Marita Inglehart. 1989. Feeling and facial efferece: implications of the vascular theory of emotion. *Psychological review* 96, 3 (1989), 395.