Gestures vs. Emojis: Comparing Non-Verbal Reaction Visualizations for Immersive Collaboration

Alexander Giovannelli, Jerald Thomas, Logan Lane, Francielly Rodrigues, and Doug A. Bowman

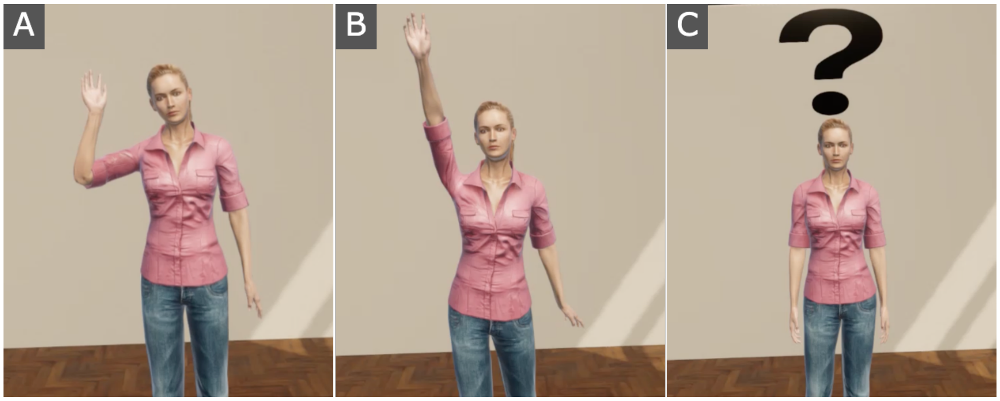

Fig. 1: View of the Question reaction performed by the female avatar across separate visualization conditions used in the experiment. A: *Natural-Question* visual reaction. B: *Exaggerated-Question* visual reaction. C: *Emoji-Question* visual reaction.

Abstract—Collaborative virtual environments afford new capabilities in telepresence applications, allowing participants to co-inhabit an environment to interact while being embodied via avatars. However, shared content within these environments often takes away the attention of collaborators from observing the non-verbal cues conveyed by their peers, resulting in less effective communication. Exaggerated gestures, abstract visuals, as well as a combination of the two, have the potential to improve the effectiveness of communication within these environments in comparison to familiar, natural non-verbal visualizations. We designed and conducted a user study where we evaluated the impact of these different non-verbal visualizations on users' identification time, understanding, and perception. We found that exaggerated gestures generally perform better than non-exaggerated gestures, abstract visuals are an effective means to convey intentional reactions, and the combination of gestures with abstract visuals provides some benefits compared to their standalone counterparts.

Index Terms—Human-computer interaction (HCI), virtual humans and avatars, telepresence, collaborative interfaces.

1 Introduction

In recent years, telepresence applications, such as Zoom [1], WebEx [2], and Microsoft Teams [3], have been leveraged by everyday consumers and enterprises to maintain the experience of an in-person collaborative environment even when users are not co-located. While these systems attempt to capture and fully display the remote participants to create the illusion of meeting in the same physical location for collaboration, they fall short of simulating face-to-face gatherings. One cause of this shortcoming is the virtual sharing of content by these systems, as the content receives greater emphasis and attention by minimizing the viewports of users or relegating the content to a secondary display separate from the virtual meeting space. This reduction in the display of collaborators reduces the ability of a given user to view the non-verbal cues, such as body gestures and facial expressions, of other collaborators [13, 16]. As an alternative, these systems include emoji reaction options (e.g., Zoom reactions [4]), however, the minimized view of the reacting participant still reduces their visibility during content sharing. As a result, users of these telepresence systems sometimes attempt to compensate for the loss of these non-verbal expressions with exaggerated vocal tones [25]. This is an issue because non-verbal cues are known to play a major role in the perceived effectiveness of collaborative interactions [6], communicating emotional states and behavioral intentions between individuals conveying and viewing them [31].

Immersive collaborative applications using virtual reality (VR) and augmented reality (AR) technologies, such as Spatial.io [5] and Virbela [6], have the potential to alleviate the issue of conveying non-verbal behaviors by embodying users via avatars. Natural gestures, such as facial expressions, gaze direction of the eyes, and body movement, can be presented by one's avatar to other users and vice versa. However, with the larger field of regard and smaller field of view provided by these systems, the saliency of non-verbal cues may still be insufficient when shared content or a greater participant pool co-inhabits a given virtual environment, making it difficult to track and understand these

- Alexander Giovannelli is with the Center for Human-Computer Interaction at Virginia Tech. E-mail: agiovannelli@vt.edu
- Jerald Thomas is with the Center for Human-Computer Interaction at Virginia Tech. E-mail: jeraldlt@vt.edu
- Logan Lane is with the Center for Human-Computer Interaction at Virginia Tech. E-mail: logantl@vt.edu
- Francielly Rodrigues is with the National Laboratory for Scientific Computing. E-mail: fmunique@posgrad.lncc.br
- Doug A. Bowman is with the Center for Human-Computer Interaction at Virginia Tech. E-mail: dbowman@vt.edu

[1] https://zoom.us/
[2] https://www.webex.com/
[3] https://www.microsoft.com/en-us/microsoft-teams/group-chat-software
[4] https://support.zoom.us/hc/en-us/articles/115001286183-Using-non-verbal-feedback-and-meeting-reactions-
[5] https://www.spatial.io/
[6] https://www.virbela.com/

subtle behaviors. Exaggerated physical gestures may alleviate this issue by increasing visibility and attracting user attention back to their collaborators from such distractors. Additionally, the use of emojis within immersive collaborative systems may enhance visibility while providing a rapid understanding of reaction intent. The conveyance of these non-verbal cues has become a topic of interest for researchers [30] in recent years, but has not closely compared methods for displaying non-verbal reactions. In this paper, we detail a study investigating avatar non-verbal reaction display methods including natural and exaggerated upper body gestures and emoji displays, as well as their combination, to examine the benefits and limitations of their impact on perception and understanding of their intended meaning.

2 Related Work

In the context of immersive technologies being used for collaboration, researchers have explored how these systems can be used for selecting the space of collaboration (i.e., co-located, remote) [21, 25], time of task fulfillment (i.e., synchronous, asynchronous) [17, 30], collaborative environment visualizations [26, 41], and cross-device collaboration [7, 15, 28]. While these works have documented nuances of avatar representations in collaborative processes, as well as means by which to evaluate communication, including performance, process, and subjective measures, they note the drawbacks of current immersive collaboration systems in properly displaying the non-verbal cues of collaborators [2, 21, 30, 32]. For this reason, we review studies of researchers spanning the influence of avatar representations and their potential effect on collaboration, as well as literature documenting artificial non-verbal communication methods in the following subsections.

2.1 Avatar representations in collaboration and interaction

Various studies have been conducted by researchers regarding the impact across quantitative and qualitative measures of visual attributes of avatars and embodied agents (i.e., computer-controlled avatars). Such visual attributes have included the display of select avatar body parts [37], as well as the stylistic appearance of the avatar [22, 36]. Yoon et al. explored the effect of avatar body visibility in collaborative processes, as well as avatar body styling on social presence [40]. The avatar body visibility had three visual conditions (i.e., head + hands, upper body, and whole body) and two appearance styles (i.e., realistic and cartoon). They found a significant effect of body part visibility on participant-reported social and co-presence, with the head + hands condition having lower values than whole body and upper body conditions. Pakanen et al. conducted a similar experiment, exploring user preferences of the degree of detail (e.g., full-body, head & arms only) as well as the degree of realism (e.g., photorealistic & cartoon) for an avatar's appearance after developing a telepresence system for immersive collaboration in augmented reality (AR) and VR [24]. Results from their study indicated for both AR and VR that participants preferred full-body avatars with photorealistic characteristics due to the human-like representation and affordances for interaction when completing a series of collaborative search and find tasks.

Researchers have also explored additional modifications of visual attributes for avatars, such as avatar scaling. For example, a study by Walker et al. [35] investigated the influence of avatar size on leader-follower relationships in a remote collaboration study. They found that, in dyadic tasks, when an avatar was scaled smaller than the other, the smaller-sized avatar commanded less attention and influence than when both avatars were of equal size. In relation to avatar scaling, Piumsomboon et al. developed a system where a remote collaborator was represented by a dynamically-scaled avatar for asymmetric and symmetric tasks, measuring the social presence and overall user experience of participants [27]. End results indicated that the "Mini-me" avatar had a significantly higher aggregate social presence, lower difficulty ratings, reduced mental effort, and completion times for experimental tasks.

Visual attribute modifications of the face and head have also been considered. Oh et al. conducted a study to determine if enhanced facial expressions (i.e., more intense smiles) would yield more positive communication outcomes compared to "normal" smiles, as well as reported perception/social presence [23]. Results offered evidence that enhancing the smile on one's avatar can lead to more positive outcomes compared to when the smile is accurately mapped on the avatar. Choudhary et al. amplify embodied facial cues by examining the impact of scaling the avatar's head to enhance non-verbal facial cue visibility [4]. They found that at various distances, slightly bigger avatar heads were more comfortable for participants in tasks relating to perceiving facial expressions and eye gaze.

Although these studies describe attribute changes to avatar physical appearance in an effort to evaluate influences on collaboration and interactions, they do not explore the potential impact of using synthetic representations for communicative cues. For this reason, we reviewed potential synthetic non-verbal display methods for further understanding the potential of these representations in communication.

2.2 Abstract visuals in affective state communication

Emotional state and producer intent via non-verbal behaviors are integral in the observational understanding of communicative exchange. There exist many methods by which these states and intents are able to be conveyed between persons, including natural and synthetic displays. In the vein of synthetic displays, multimodal metaphors as described by Forceville and Aparisi are a means by which metaphors in language can be applied in thought and action with equivalent cognitive mapping as their verbal counterparts [12]. Such metaphors have been identified and researched from pictorial runes in Japanese comics by Shinohara and Matsunaka, who identified affective representations of anger, surprise, joy, disappointment, and love [33]. However, they do note that the natural phenomena displayed by the pictorial runes may be culture-specific and require further examination. Eerden also discusses the usage of pictorial runes and multimodal metaphors in his work, describing how they are capable of being used to depict abstract concepts which would be otherwise difficult to depict literally [8]. He further explains the potential of using such runes for emotions, stating that cartoonists often leverage pictorial runes to show these states via smoke, jagged lines, spirals, and color selection.

Color has also been leveraged as a means to communicate affective state and behavioral intent [18]. Bartram et al. investigated the relations of color properties (e.g., hue, chroma, lightness), as well as composition (e.g. color frequency and clustering) in an effort to relate affective impressions to simple visualizations [1]. From three separate studies they conducted, they found that composition and properties of color achieve affective expressiveness in visualizations when manipulated according to the intent desired by a producer. This was also explored by Şemsioğlu et al. in a tool they created to provide affective augmentations using colored shapes surrounding human silhouettes for designers [5]. It provided predefined color points on a diagram representing Russell's circumplex model of affect, similar to work from Valente et al. that visualized system inferred the emotional state of users to surround them with colored auras based on the same model [34].

Emojis and emoticons are popular ways to communicate over digital media and convey affective state of the producer. Wiseman and Gould examined the usage of emojis as their own "ubiquitous language", exploring the highly personalized and purposefully secretive ways emojis are used in conversation between various relations (e.g., partner, friend, family) [38]. They referred to this as "emoji affordance" and conducted further exploration into the underlying phenomena that lead to the repurposing of emojis. Fischer and Herbert add upon emoji meaning analysis explaining that, compared to other mediums of emotion or intent conveyance, emojis are ambiguous and do not symbolize discrete affective states [11]. They conducted a study specifically investigating emoji ratings regarding affective state, comparing the results with emoticons and human faces. From their study, they found emojis elicit the highest arousal from participants, detailing that discrete emotion was also best recognized in emoji, contrary to their initial claim. Although these studies provide insight into affective state and intent in a producer-observer relationship, they do not study these synthetic displays used with real-time avatar counterparts.

From the existing literature discussed in this section, work regarding the expressivity of gestures in the form of natural and exaggerated movements has been limited to the facial features of avatars. The extent to which exaggerated body gestures conducted by an avatar influence a user's ability to identify and understand their intent, as well as their observed perception, has not been researched to our knowledge. Additionally, while research has explored the usage of abstract visuals in the form of emojis and other semiotics compared to pictures of real people or used in the context of social VR, their impact and potential implications in augmenting or replacing gestural behaviors exhibited by an in real-time avatar has not been studied. Our study intends to address these gaps in research by having participants complete mock collaborative tasks while an avatar performs non-verbal reactions with different visualizations, including natural/exaggerated gestures and emojis.

3 Experiment

The experiment detailed in this section was designed with the goal of comparing the effectiveness of different methods of communicating non-verbal reactions via an avatar. Specifically, we wanted to understand the trade-offs in effectiveness between natural gestures and the alternative methods of exaggerated gestures and abstract visualizations. Our experiment aimed to answer three research questions.

RQ1. How does the expressivity of an avatar's gestural reactive behaviors influence an observer's ability to identify a reaction that is occurring and understand which reaction is being conveyed?

Past study results suggest that exaggerated avatar features are more positively perceived and better understood by observers [4, 23]. However, the challenge lies in exaggerating entire avatar gestures while maintaining a sense of naturalness. We want to determine whether increasing the expressiveness of avatar gestures through exaggeration yields a positive effect and, if so, to what extent. Based on previous research, we had the following hypotheses for RQ1:

- *H1.1* Exaggerated gestures will be identified faster than non-exaggerated gestures.
- *H1.2* Exaggerated gestures will be identified with more confidence than non-exaggerated gestures.
- *H1.3* Exaggerated gestures will be preferred by users over non-exaggerated gestures.

RQ2. How does using abstract visuals instead of gestural reactive behaviors influence an observer's ability to identify a reaction that is occurring and understand which reaction is being conveyed?

Past studies suggest that abstract visuals, such as emojis, are more easily recognized and understood by observers than physical non-verbal cues [11, 19]. However, those studies related to abstract visuals in direct comparisons of facial images. For this reason, we want to determine how abstract visuals in the form of overhead emojis are perceived and if they are understood more effectively in immersive VR using avatars. Based on previous research, we had the following hypotheses for RQ2:

- *H2.1* Abstract visuals will be identified faster than gestures.
- *H2.2* Abstract visuals will be identified with more confidence than gestures.
- *H2.3* Abstract visuals will be preferred by users over gestures.

RQ3. Compared to either abstract visuals or gestural reactive behaviors alone, how does the combination of the two influence an observer's ability to identify a reaction that is occurring and understand which reaction is being conveyed?

We expect the combination of gestures and abstract visuals to result in the most performant and preferred means by participants to identify and understand non-verbal cues, as it includes the natural physical movement with the reinforced understanding through the emoji. For this reason, we had the following hypotheses for RQ3:

- *H3.1* The combination of gestures and abstract visuals will be identified faster than their standalone equivalents.
- *H3.2* The combination of gestures and abstract visuals will be identified with more confidence than their standalone equivalents.
- *H3.3* The combination of gestures and abstract visuals will be preferred by users over their standalone equivalents.

3.1 Conditions

In the experiment, three reaction conditions were used: Agree, Disagree, and Question. These reactions were conveyed using five visualization conditions: NATURAL, EXAGGERATED, EMOJI, NATURAL & EMOJI, and EXAGGERATED & EMOJI. Two avatars, one male and one female (shown in Fig. 1), were used to perform the reactions for each of the visualizations. The NATURAL visualization consisted of subtle gestural animations by the avatars: lightly nodding their head up and down for the Agree reaction, shaking their head side-to-side for the Disagree reaction, and raising their hand to shoulder height for the Question reaction. The EXAGGERATED visualization enhanced the NATURAL animations for the avatars, having a faster or larger range of motion gestures to the previously described reactions. The EMOJI visualization displayed semiotic visuals above the avatars' heads in place of the gestural animations: a check mark for the Agree reaction, an 'X' for the Disagree, and a question mark for the Question reaction. The NATURAL & EMOJI and EXAGGERATED & EMOJI visualizations were a combination of the gestural animations with the simultaneous display of the semiotic visuals above the avatars' heads. The NATURAL, EXAGGERATED, and EMOJI visualizations are shown for the Question reaction in Fig. 1.

3.2 Tasks

Participants completed three separate tasks as part of the study phase of the experiment within our VE. The first task was referred to as the **Open Classification** task, which consisted of 18 trials, where each trial showed a unique combination of avatar gender (2), basic visualization (3), and reaction (3). This task was included in order to determine whether the different visualizations of each reaction were intuitively understandable on their own, without prior explanation. A single trial had participants observe an avatar in our VE as it performed a non-verbal reaction for one of three visualizations: NATURAL, EXAGGERATED, and EMOJI. After the reaction was performed by the avatar, the participant was prompted to describe what they believed the avatar was trying to convey to them out loud using a one-word descriptor. The trial was completed once the investigator transcribed the word provided by the participant. The participant would then move on to the next trial, consisting of the alternate avatar (e.g., if the previously displayed visual reaction was made by the male avatar, the female avatar would take its place) conveying another reaction with a unique visualization display condition. The order by which the visualizations, reactions, and avatars performed the non-verbal displays in the task was fixed for all participants.

The second task was the **Training** task, which consisted of six trials. The purpose of this task was to train the participant to recognize the selected reaction animations and further validate that they were distinguishable for each visualization. For a given trial, three copies of the same gender avatar were displayed in the VE equidistant in front of the participant. Simultaneously, all three avatars performed the same reaction, each using a specific visualization: NATURAL, EXAGGERATED, and EMOJI. The participant reported which avatar performed a particular visualization as prompted by the investigator. In cases where the participant misreported the visualization, they were informed by the investigator which avatar correctly acted out the prompted visualization, followed by a replay of the reaction. The investigator would then record that the identification was failed and the trial would be re-queued for additional training after all other reaction training trials. This marked the end of a trial, after which another trial with the alternate avatar or new reaction would be used. The order in which reactions, visualizations, and performing avatars were displayed, as well as prompts by the investigator for the visualization condition, were uniform across participants.

The third, and primary, task was called the **Detection** task, which consisted of ten trials in the VE shown in Fig. 2. This task was meant to simulate collaboration, where participants would read from a script placed in front of them within the VE. The script was meant to take participant attention away from the avatar, similar to shared documents taking attention away from collaborators in telepresence applications. During a trial, the avatar exhibited reactions associated with a specific visualization. These reactions were displayed once each while the participant read a script to the avatar. Using a controller, the participant reported whenever they believed they observed an Agree, Disagree, or Question reaction being performed by the avatar. The order of visualizations was counterbalanced using a Latin square, with the avatar display order alternating between participants. Reactions performed by the avatar were triggered at specific points in the script, corresponding to certain words known only by the investigator. Additionally, distracting gestures such as yawning and scratching of the head were made by the avatar for all gesture-based conditions (i.e., all visualizations except Emoji), to simulate physical gestures that are made without intent to convey a reaction. A trial ended once the participant completed reading the script.

Fig. 2: Virtual environment during the **Detection** task.

3.3 Design

This was a within-subjects study that had each participant complete the tasks as described in Section 3.2. Independent variables in the experiment were the visualization, avatar gender, and reaction conditions. Dependent variables measured during the experiment were task-specific. For the Open Classification task, the dependent variable was the participant's one-word descriptor. In the Training task, incorrect identifications made by the participant when identifying visualizations were recorded by the investigator. The intent of the task was to serve as a means to build participant understanding of the differences in the reactions between varying visualizations. Finally, in the Detection task, dependent variables included the time taken by the participant to report a reaction being performed by an avatar for a given trial, the self-reported confidence of their reaction identification, the percentage of time spent by the participant gazing at the avatar during each trial, and the participant's subjective ratings of the visualizations on a between-block questionnaire. The between-block questionnaire had the participant provide Likert scale ratings to the prompt of "The reactions of the avatar were..." for the following pairings: *Unnatural(1)-Natural(7)*, *Confusing(1)-Understandable(7)*, *Unprofessional(1)-Professional(7)*, *Misleading(1)-Trustworthy(7)*, *Unnoticeable(1)-Noticeable(7)*, and *Informal(1)-Formal(7)*. Open responses made by the participant during the Open Classification task were transcribed by the investigator. Report and gaze times, as well as identification confidence values, were recorded automatically by the system as part of the Detection Task, with between-block questionnaires completed by the participant after each trial.

3.4 Apparatus

Participants were seated and wore a Varjo Aero [7] head-worn display (HWD) for the experiment. The HWD has a 115-degree horizontal and 83-degree vertical (38 degrees upward and 45 degrees downward) field of view and a resolution of 2880x2720 px per eye with a refresh rate of 90 Hz. The HWD has eye tracking at 200 Hz with sub-degree accuracy. Four HTC Vive Base Stations were used to track the participant's position within the virtual environment and to establish a stationary boundary. Valve Index controllers were used by the investigator and participant for controller-based actions within the experiment.

3.5 Environment

The virtual environment used in the experiment was a box-shaped room with dimensions 6m x 4m x 3m (length x width x height). The avatars were positioned 2.75m in front of the participant with uniform height. As part of the reading task, the script stand was set to approximately 0.4m away from the participant's head, a reading distance considered to be normal in adults with perfect vision [3].

3.6 Avatars

The avatars used in the experiment were from Microsoft Rocketbox, a library that consists of 115 fully rigged, high-definition characters and avatars [14]. Agree and Disagree reactions for the NATURAL and EXAGGERATED visualizations made by the avatar used animations provided with the Microsoft Rocketbox library. The Question reaction for the NATURAL and EXAGGERATED visualizations made by the avatar used Mixamo [8], an auto-rigging software for easy avatar rigging and animation. EMOJI visualization images for all reaction conditions were captured from online resources. They were modified to use the same color (i.e., black) and size, and were positioned identically above each avatar's head.

3.7 Participants

Thirty participants were recruited from undergraduate computer science courses and received volunteer credit for their participation. All participants self-reported that they met the following eligibility criteria from a pre-screening survey: they were 18 years of age or older, had fluent proficiency in English, had perfect vision or used corrective contact lenses, and did not suffer from any vision deficiencies. The ages of the participants ranged between 18 to 22 years (M=20.2, SD=0.81). 22 participants were male and eight were female. Three participants reported they had not used VR previously, 12 had used VR once or twice, ten had used VR three to ten times, and five had used VR more than ten times. 28 participants were right-hand dominant, one was ambidextrous, and one was left-hand dominant. The study was approved by our local Institutional Review Board.

3.8 Procedure

Each participant completed three phases as part of the experiment: pre-study, study, and post-study. In the pre-study phase, the participant was given an informed consent document to read and sign. The participant then filled out a Simulator Sickness Questionnaire (SSQ) [20]. Finally, the participant was shown the HWD and instructed on how to adjust the head straps and locate the menu buttons for eye-tracking calibration before moving to the study phase.

In the study phase, the participant completed the three tasks described in Section 3.2. During the Open Classification task, the participant put on the HWD while seated and eye-tracking calibration was performed. On confirmation of the eye-tracking calibration by the investigator, height calibration was performed such that all participants had the same head position relative to the avatar for the duration of the experiment. After these calibration processes, the participant would perform the trials as part of the task. For each trial, the participant was able to request the replay of the avatar's reaction to determine their one-word descriptor. Once the participant completed all 18 trials, they

[7] https://varjo.com/products/aero/
[8] https://www.mixamo.com/

were instructed to remove the HWD and were given a break prior to the Training task.

During the Training task, the participant was introduced to the Valve Index controller. It was explained to the participant that the controller would be used to report avatar reactions with a built-in survey during the task. The participant then put on the HWD while seated and eye-tracking calibration was performed. On investigator confirmation of eye-tracking calibration, the participant was given the Valve Index controller and instructed on how to perform built-in reaction report questionnaires, shown in Fig. 3. Once the participant indicated they were familiar with the controls to perform the reaction report, the investigator started the task's trials. For each trial, the investigator informed the participant as to which reaction condition was performed by the avatars (e.g., "The reaction shown was the agree reaction for the female avatar."), prior to prompting the participant to identify which avatar conveyed a specific visualization (e.g., "Can you identify which avatar conveyed the exaggerated reaction?"). The participant was able to request having the avatars repeat the reaction as necessary to determine the avatar whose visualization matched the prompt. In the case the participant identified the visualization incorrectly, the investigator identified the avatar that conducted the prompted visualization and re-queued the reaction to re-train the participant later on, recording the participant had failed to differentiate the visualization conducted by the avatar. Otherwise, the participant was informed their decision was correct and was instructed to report the avatar's reaction and their identification confidence using the built-in identification questionnaire, resulting in the end of the trial. This process was repeated until all visualizations were correctly identified by the participant. The participant was then instructed to remove the HWD and was given a break prior to the Detection task.

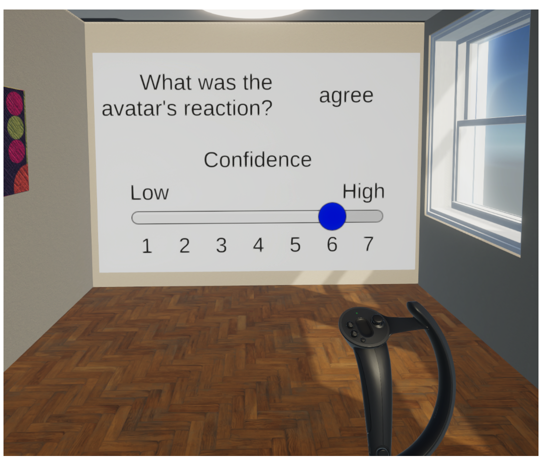

Fig. 3: Built-in reaction identification questionnaire.

The Detection task consisted of 10 trials, each of which included three reactions to be detected and identified. Prior to each trial, the participant put on the HWD while seated and eye-tracking calibration was performed. On investigator confirmation of eye-tracking calibration, script height calibration was performed such that the script was approximately 0.4m away from the participant's head while facing the avatar in the VE. The participant was then given the Valve Index controller and informed the reporting controls were identical to the prior task. Before the trial started, the participant was asked whether they were able to clearly view the script text before reading, such that script height calibration could be redone if necessary. The participant was then informed as to which visualization would be used to convey reactions by the avatar (e.g., "The visualization under test in this trial is the NATURAL & EMOJI visualization."). Additionally, the investigator informed the participant to open the identification questionnaire (Fig. 3) using the controller as soon as they believed they witnessed a valid reaction. Participants then completed the trial as described in Section 3.2. At the end of each trial, the participant was instructed by the investigator to remove the HWD and complete a between-block questionnaire regarding the visualization they experienced. After completing all trials for the task, the study phase was completed and the participant moved on to the post-study phase.

In the post-study phase, the participant filled out another SSQ. After completing the SSQ, a survey was given to the participant for ranking the visualizations for (1) effectiveness in conveying reactions and (2) personal preference in conveying the reactions. Once the participant completed the ranking survey, a verbal interview was conducted in which the investigator asked the participant for the reasoning behind their rankings for each condition, as well as solicited any feedback regarding differences in reaction effectiveness across visualizations. After the interview, the participant completed a background questionnaire, which recorded information regarding the participant's gender, age, dominant hand, and experience using VR. This marked the end of the post-study phase and experiment.

4 RESULTS

We conducted a series of analyses on the dependent variables (described in Section 3.3) collected during our experiment to test our hypotheses and explore the benefits and drawbacks of non-verbal visualizations across reactions. Prior to our analysis, we removed any trials in which the participant did not correctly identify the avatar's reaction (e.g., the participant reported the avatar made a "question" reaction when the avatar in fact made an "agree" reaction, the participant reported the avatar made a reaction when there was not a reaction, etc.). There were not enough incorrect trials (i.e., missed reactions, false positives, and incorrect identifications) to draw any meaningful conclusions about the effects of the independent variables on these measures.

An initial analysis revealed a significant interaction effect of avatar gender with the visualization and reaction conditions. Our design included avatar gender as an explicit independent variable with the intention of evaluating its possible effects on viewing and understanding reactions, so long as the gender of participants was balanced and no differences in expressivity between NATURAL and EXAGGERATED animations were identified. However, the participant gender was not balanced, and based on records for each participant from the Training task regarding avatar visualization identifications, the animations were not equally identifiable between avatars. Specifically, 33% of the participants misreported the *Natural-Agree* reaction and 23% misreported the *Natural-Disagree* reaction for the male avatar, as opposed to 3% for the *Natural-Disagree* of the female avatar (due to the organization of the task, we do not have comparative data for the *Natural-Agree* of the female avatar). Participants further noted this difficulty in distinguishing avatar animations in post-study interviews, with P1 stating that the *Exaggerated-Disagree* reaction for the male avatar "felt slightly more excessive than the regular one, but only slightly", P7 saying that the female avatar for the EXAGGERATED visual reactions compared to the male avatar "was a bit more subtle", and P21 describing that the "male natural seemed on par with the exaggerated." As a result, we cannot assess if differences in reported measures were caused by the avatar gender or these variances in animations. For this reason, we removed avatar gender as a factor in our analysis and report our results based on visualization and reaction conditions using the female avatar data only. Lastly, SSQ scores are not reported, as pre- and post-experiment results did not vary on review.

4.1 Open Classification

One-word responses from participants were transcribed as part of the Open Classification task. These participant responses were compared to the intended reaction descriptor (i.e., Agree, Disagree, Question) and were considered a match if synonymous. For the NATURAL visualization, 87% of participants correctly identified the Agree reaction, 90% the Disagree reaction, and 10% the Question reaction. In the EXAGGERATED visualization, 90% of participants correctly identified the Agree reaction, 90% the Disagree reaction, and 17% the Question reaction. Finally, during the EMOJI visualization, 93% of participants correctly identified the Agree reaction, 90% the Disagree reaction, and 47% the Question reaction. The Question reaction for both gesture conditions was most commonly identified as a greeting, with 90% in NATURAL and 47% in EXAGGERATED reporting this.

4.2 Reaction Report Time

A Kolmogrov-Smirnoff goodness of fit test was conducted for the reaction report time across visualizations for all reaction conditions. The data was not normally distributed, so we applied an Aligned Rank Transform (ART) before using a Repeated-Measures Analysis of Variance (RM-ANOVA) for factorial analysis [39]. The ART RM-ANOVA indicated a significant effect of the visualization ($F(4,389.20) = 74.450, p < 0.01$), reaction ($F(2,389.20) = 14.862, p < 0.01$), and Visualization×Reaction interaction ($F(8,389.20) = 16.123, p < 0.01$) on reaction report time. We then used the ART-C algorithm for multifactor post-hoc contrast tests on the Visualization×Reaction interaction with Bonferroni adjustments on all pairwise comparisons [10].

For the Agree reaction, we found that the NATURAL visualization took significantly longer to report compared to EMOJI ($p < 0.01$), NATURAL & EMOJI ($p < 0.01$), and EXAGGERATED & EMOJI ($p < 0.01$) visualizations. Additionally for the Agree reaction, we found that the EXAGGERATED visualization took significantly longer to report compared to the EMOJI ($p < 0.01$), NATURAL & EMOJI ($p < 0.01$), and EXAGGERATED & EMOJI ($p < 0.01$) visualizations. For the Disagree reaction, we found that the NATURAL visualization took significantly longer to report compared to EXAGGERATED ($p = 0.048$), EMOJI ($p < 0.01$), NATURAL & EMOJI ($p < 0.01$), and EXAGGERATED & EMOJI ($p < 0.01$) visualizations. Additionally for the Disagree reaction, we found that the EXAGGERATED visualization took significantly longer to report compared to the EMOJI ($p < 0.01$), NATURAL & EMOJI ($p < 0.01$), and EXAGGERATED & EMOJI ($p < 0.01$) visualizations. Median and IQR values for each visualization and reaction are given in Table 1.

4.3 Identification Confidence

As the identification confidence data is nonparametric (i.e., 7-point Likert scale reported values), we used an ART RM-ANOVA for factorial analysis. The analysis indicated a significant effect of the visualization ($F(4,389.60) = 41.834, p < 0.01$), reaction ($F(2,389.60) = 55.068, p < 0.01$), and Visualization×Reaction interaction ($F(8,389.60) = 12.991, p < 0.01$) on identification confidence. We then used the ART-C algorithm for multifactor post-hoc contrast tests on the Visualization×Reaction interaction with Bonferroni adjustments on all pairwise comparisons.

For the Agree reaction, we found that the NATURAL visualization had significantly lower identification confidence values compared to EXAGGERATED ($p < 0.01$), EMOJI ($p < 0.01$), NATURAL & EMOJI ($p < 0.01$), and EXAGGERATED & EMOJI ($p < 0.01$) visualizations. For the Disagree reaction, we found that the NATURAL visualization had significantly lower identification confidence values compared to EXAGGERATED ($p < 0.01$), EMOJI ($p < 0.01$), NATURAL & EMOJI ($p < 0.01$), and EXAGGERATED & EMOJI ($p < 0.01$) visualizations. For the Question reaction, we found that the NATURAL visualization had a significantly lower identification confidence value compared to the EXAGGERATED & EMOJI ($p = 0.014$) visualization. Median and IQR values for each visualization and reaction are given in Table 1.

4.4 Gaze Percentage

Fig. 4c shows the mean percentage of time spent gazing at the avatar across visualizations. A Kolmogrov-Smirnoff goodness of fit test found that the data was not normally distributed, so we applied a square root transform before performing a one-way RM-ANOVA. The RM-ANOVA indicated significance between the visualization conditions. However, sphericity was violated according to Mauchly's Test ($\chi^2(df = 9, N = 30) = 24.354, p = 0.004$), so we used Greenhouse-Geisser adjustments and found significance ($F(9,30) = 23.288, p < 0.01$). Pairwise comparisons were performed between visualizations with Bonferroni adjustments.

We found that the NATURAL visualization had significantly higher gaze percentages compared to EXAGGERATED ($p = 0.024$), EMOJI ($p < 0.01$), NATURAL & EMOJI ($p < 0.01$), and EXAGGERATED & EMOJI ($p < 0.01$) visualizations. We found that the EXAGGERATED visualization had significantly higher gaze percentages compared to EMOJI ($p < 0.01$), NATURAL & EMOJI ($p = 0.023$), and EXAGGERATED & EMOJI ($p = 0.024$) visualizations. We found that the EXAGGERATED & EMOJI visualization had a significantly higher gaze percentage compared to EMOJI ($p < 0.01$) visualization. Median and IQR values for each visualization are given in Table 1.

4.5 Between-block Questionnaire

For each scale in the between-block questionnaire, Friedman tests were conducted to determine if there was a significant effect of visualization. No significant effect was discovered between visualizations for *Formal* ($\chi^2(df = 4, N = 30) = 3.209, p = 0.524$) and *Professional* ($\chi^2(df = 4, N = 30) = 7.611, p = 0.107$) scales. For the remaining scales, the following significant effects were found: *Natural* ($\chi^2(df = 4, N = 30) = 47.831, p < 0.01$), *Noticeable* ($\chi^2(df = 4, N = 30) = 65.407, p < 0.01$), *Trustworthy* ($\chi^2(df = 4, N = 30) = 18.892, p < 0.01$), and *Understandable* ($\chi^2(df = 4, N = 30) = 40.979, p < 0.01$). For each of these scales, paired Wilcoxon signed-rank tests with Bonferroni adjustments were performed between the visualizations. The resulting significant pairs from these tests are in Table 2. Median and IQR values for each visualization and reaction are given in Table 1.

5 DISCUSSION

5.1 Open Classification

The Open Classification task was meant to confirm the participant's approximate understanding of the intended meaning of the gestures and emojis without prior training. From the one-word responses, we found that participants were generally able to properly understand the agree and disagree reactions for NATURAL, EXAGGERATED, and EMOJI visualizations. However, the participants had difficulty identifying the Question reaction, identifying it primarily as a greeting for the NATURAL and EXAGGERATED visualizations, and as a sign of confusion for the EMOJI visualization. After the participant was trained in the Training task, they understood the intended definition of the Question reaction. In the post-study interview, P15 discussed this difference between the Agree and Disagree reactions compared to that of the Question reaction, stating they felt "raising your hand is not a universal sign of question, but... agree and disagree are pretty universal and easy to understand." Thus, with the additional context, the gestures and emojis used throughout the duration of the experiment were understood by participants.

5.2 Expressivity of Gesture

We hypothesized that more expressive gestures would be identified faster by participants than less expressive gestures (*H1.1*). From measured reaction report times and trial gaze summary data, our results partially support *H1.1*. Although all reactions had lower reaction report times for the EXAGGERATED visualization than that of the NATURAL visualization, the only significant difference was for the Disagree reaction. However, the percentage of time spent by the participants gazing at the avatar during a given trial was significantly greater for the NATURAL visualization. This suggests that, while participants were able to identify the avatar's reactions similarly quickly between the two visualizations, less expressive gestures required the participant to pay more attention to the avatar than the primary task of reading. Several participants discussed this issue during open interviews, with P8 describing that NATURAL visualization of reactions was "difficult to see sometimes or distinguish when you're not focused on the actual person," P16 saying it was "harder to pick up because you're reading and then you see a very slight thing," and P27 stating "While I was reading, if I wasn't focused (on the avatar), I would have missed it." Comparatively, the expressive gestures conveyed by the avatar for the EXAGGERATED visualization were better at drawing the participants' attention from the script, with P11 stating that "it grabs my attention... but I wasn't directly looking at it," P12 saying "Your eye is drawn to the movement away from the script," and P20 describing exaggerated gestures as "easier to see cause more movement in my peripheral vision."

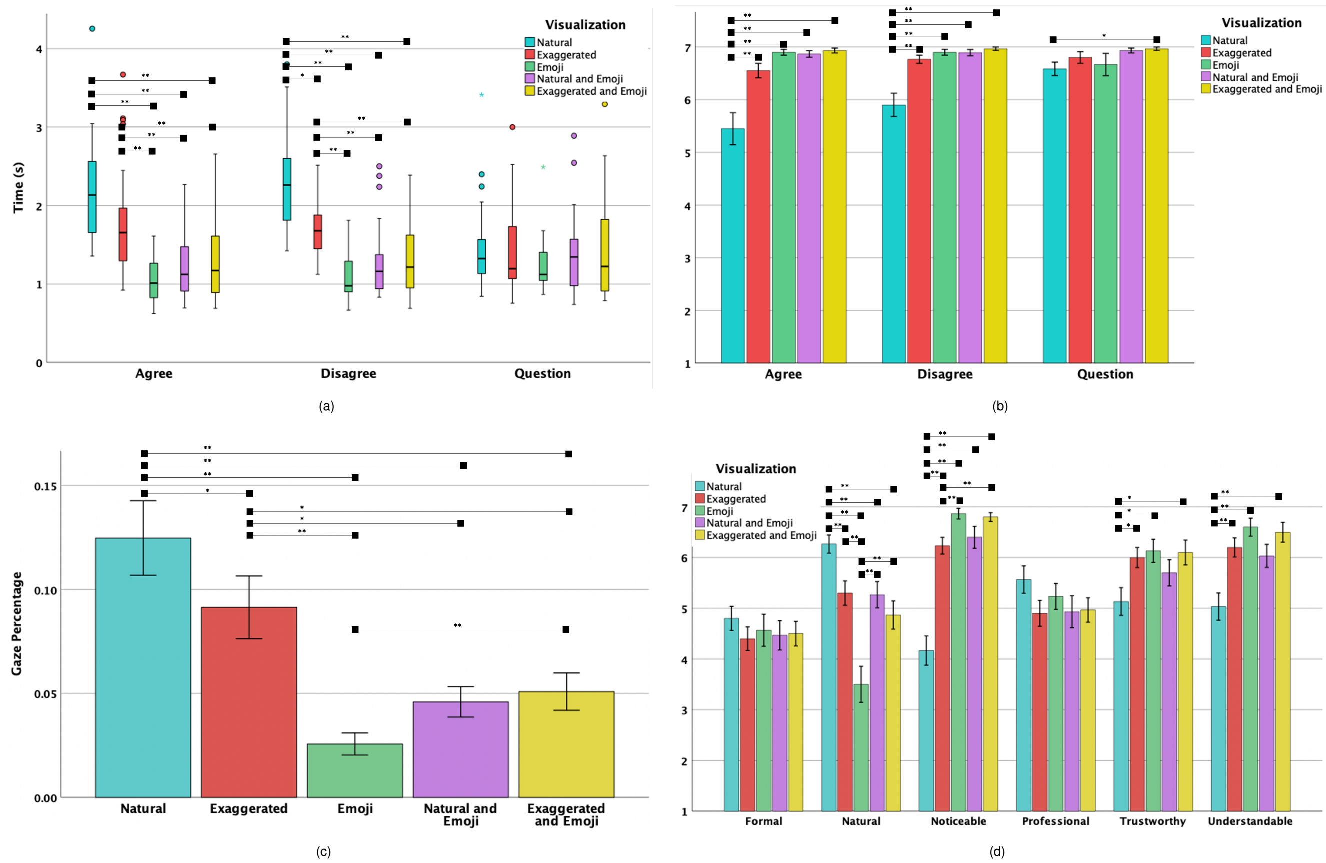

Fig. 4: Plots of identification and perception measures. Pairs that are significantly different are marked with * when $p \leq 0.05$ and ** when $p \leq 0.01$. Box plot whiskers are the spread of the data (without outliers). Bar chart whiskers are the $\pm S.E.$ spread of the data. (a): Time taken to report reactions. (b): Reported identification confidence. (c): Percentage of time spent gazing at the avatar. (d): Between-block questionnaire results.

Additionally, we hypothesized that expressive gestures would be identified with more confidence than less expressive gestures (*H1.2*). Based on participant-reported identification confidence data and post-study discussion, our results support *H1.2*, albeit with an interesting exception. The reported identification confidence increased with more expressive gestures conveyed by the avatar, with analysis revealing significantly greater confidence values for the EXAGGERATED visualization of the Agree and Disagree reactions. Although the Question reaction identification confidence value did increase for the EXAGGERATED visualization, this difference was not significant. P3 noted this, stating "I think the questioning were both very obvious because they raise your hand in a particular way, so there's no difference there." Participants further noted the benefits of the EXAGGERATED visualizations, with P5 saying that more expressive movement helped when reading since "something has to move in the corner of my eye for like me to notice it" and P13 stating "the exaggerated ones had a lot more body language, while natural, had less of it, so that's why I thought the exaggerated was a lot easier to distinct."

Finally, we hypothesized that more expressive gestures would be preferred by participants over less expressive gestures (*H1.3*). From participant visualization rankings, between-block questionnaire values, and post-study interview discussions, our results partially support *H1.3*. Visualization rankings conducted in the post-study had 57% of participants identify EXAGGERATED as their preferred visualization compared to NATURAL. Between-block questionnaire results showed a significant difference between Professional, Natural, Noticeable, and Trustworthy values reported between the NATURAL and EXAGGERATED visualizations. The NATURAL visualization was rated as more natural. Several participants described their rationale behind these scores, with P1 describing the NATURAL visualization as "more genuine and more human" P9 saying "the natural reactions made the most sense because that's just how we communicate," and P18 stating the NATURAL visualization is "comforting, but traditional" and "less jarring." The greater movement by the avatar in the EXAGGERATED visualization also negatively impacted it for the Natural measure, with P23 describing the reactions as being "crazy" and P8 calling them "more cartoony." However, the EXAGGERATED visualization scored higher for noticeability, trustworthiness, and understandability. P12 stated that the avatar's exaggerated reactions made them "certain of their intentions," P6 explicitly said "very easy to notice," and P7 described the exaggerated reactions as "definitely easier to understand."

Overall, we found that when participants focused their attention on a content-related task, exaggerated gestures made by an avatar had a positive effect on the participant's ability to perceive and understand their intended meaning compared to their more natural counterparts. Exaggerated behaviors empower participants to pay more attention to their work, while not sacrificing their ability to notice an avatar's reactions and confidently identify the intended meaning of the reactions. However, although these exaggerated behaviors may be preferred by most participants, they can be perceived as informal and unprofessional, making the setting and context of collaboration important considerations.

5.3 Gestures versus Abstract Visuals

When comparing abstract visuals with gestures, we decided to select the more performant visualization condition. This was determined to be EXAGGERATED after analysis in Section 4. Thus the following discussion compares the EXAGGERATED and EMOJI visualizations.

We hypothesized that abstract visuals would be identified faster by participants than gestures (*H2.1*). Based on participant reaction report times and trial gaze summary data, our results support *H2.1*. All reaction report times decreased for the EMOJI visualization compared to the EXAGGERATED visualization, with significantly lower report times for Agree and Disagree reactions. Additionally, the percentage of time

		Natural		Exaggerated		Emoji		Natural & Emoji		Exaggerated & Emoji	
		Med	*IQR*	*Med*	*IQR*	*Med*	*IQR*	*Med*	*IQR*	*Med*	*IQR*
Report Time	Agree	2.13	0.92	1.66	0.72	1.01	0.45	1.12	0.57	1.17	0.73
	Disagree	2.26	0.80	1.68	0.47	0.98	0.41	1.16	0.45	1.22	0.75
	Question	1.32	0.51	1.19	0.70	1.12	0.37	1.34	0.61	1.22	0.93
Confidence	Agree	6.0	1.0	7.0	1.0	7.0	0.0	7.0	0.0	7.0	0.0
	Disagree	6.0	2.0	7.0	0.3	7.0	0.0	7.0	0.0	7.0	0.0
	Question	7.0	1.0	7.0	0.0	7.0	0.0	7.0	0.0	7.0	0.0
Gaze Percentage		0.10	0.15	0.06	0.11	0.02	0.03	0.04	0.05	0.04	0.05
Questionnaire	Formal	5.0	2.0	4.0	2.0	5.0	3.0	4.50	2.0	4.0	2.0
	Natural	7.0	1.0	5.0	1.0	3.0	3.0	5.0	5.0	5.0	2.0
	Noticeable	4.0	2.0	6.5	1.0	7.0	0.0	7.0	1.0	7.0	0.0
	Professional	6.0	2.0	5.0	2.0	5.0	1.0	5.0	4.0	5.0	2.0
	Trustworthy	5.0	2.0	6.0	2.0	7.0	1.0	6.0	2.0	7.0	1.0
	Understandable	5.0	2.0	6.5	1.0	7.0	0.0	6.5	2.0	7.0	1.0

Table 1: All median (*Med*) and interquartile range (*IQR*) values for each visualization across study measures.

Scale	Vis. 1	Vis. 2	Z	p-value
Natural	Exaggerated	Natural	-3.403	p <0.01
	Emoji	Natural	-4.376	p <0.01
	Natural & Emoji	Natural	-3.583	p <0.01
	Exaggerated & Emoji	Natural	-3.744	p <0.01
	Emoji	Exaggerated	-3.425	p <0.01
	Natural & Emoji	Emoji	-3.581	p <0.01
	Exaggerated & Emoji	Emoji	-3.619	p <0.01
Noticeable	Exaggerated	Natural	-4.129	p <0.01
	Emoji	Natural	-4.488	p <0.01
	Natural & Emoji	Natural	-4.201	p <0.01
	Exaggerated & Emoji	Natural	-4.475	p <0.01
	Emoji	Exaggerated	-3.211	p = 0.01
	Exaggerated & Emoji	Exaggerated	-3.220	p = 0.01
Trustworthy	Exaggerated	Natural	-3.052	p = 0.02
	Emoji	Natural	-2.853	p = 0.04
	Exaggerated & Emoji	Natural	-2.786	p = 0.05
Understandable	Exaggerated	Natural	-3.432	p <0.01
	Emoji	Natural	-4.155	p <0.01
	Exaggerated & Emoji	Natural	-3.850	p <0.01

Table 2: Significant pairings determined by the Wilcoxon signed-rank tests for the between-block questionnaire scales.

spent by participants gazing at the avatar was significantly lower for the EMOJI trials than for EXAGGERATED trials. Participants discussed the benefits of EMOJI visualizations in reporting reactions during post-study discussions. P24 felt they could "immediately see something pop up above the (avatar's) head," P6 stated that emoji were "very easy to see," and P30 described that "the emojis showed instantly what they (the avatar) were portraying without having to wait to see how their body would move." The Question reaction, however, was not reported significantly faster by participants for the EMOJI visualization. We attribute this to the difference in the range of motion of the animation. While EXAGGERATED Agree and Disagree reaction animations span the upper body, with avatar head shaking motions and hand movements, the Question reaction is a full-body gesture that has the avatar step forward while extending their arm fully upward. Participants highlighted this in post-study discussions, with P5 commenting "with the question exaggerated, I was confused why it was moving forward," P9 described "that little step forward that the avatar would do made the whole motion very noticeable," and P26 stated that the exaggerated question animation was "very dynamic."

We also hypothesized that abstract visuals would be identified with more confidence than gestures (*H2.2*). From the participant identification confidence values, our results do not support *H2.2*. The reported identification confidence by participants had no significant differences between EMOJI and EXAGGERATED gestures conveyed by the avatar across reactions. However, when discussing the visual reactions, participants described how EMOJI visualizations aided them in their identifications. P13 described the emojis as "very distinctive" and that "when I saw that picture above their head, I knew exactly what that reaction was," as well as P1 stated that "emoji are very self-evident in what they are trying to convey" along with P15 saying that emoji "really help people understand what you're conveying."

Our last hypothesis was that abstract visuals would be preferred by participants over gestures. Based on participant visualization rankings, between-block questionnaire values, and post-study interview discussions, our results partially support *H2.3*. Visualization rankings conducted in the post-study had 57% of participants identify emoji as their preferred visualization compared to EXAGGERATED. With the exception of a significantly lower natural value, the EMOJI visualization had higher reported values than the EXAGGERATED visualization in the between-block questionnaire. In particular, emojis were rated as significantly more noticeable. Participants described their preference toward EMOJI visual reactions in the post-study interview, with P3 explaining that emoji are "more effective than the just exaggerated," P12 stating that emoji aided them in "being certain of what people mean," and P25 describing that an emoji "shows what the reaction is supposed to be, is obviously very noticeable, but also very clear in what the message is" and that "facial expressions can be interpreted different ways." However, participants also described that the EMOJI visualization was not always preferable, with P9 explaining that "having absolutely no human bodily reaction, just an emoji pop up above your head is in no way natural" and that it felt as if they were having a "very one-sided conversation." P21 also stated they "wouldn't care to see emojis alone" and P8 described emoji as "not very fun or engaging."

Overall, we found that emojis provided an immediate means by which participants were able to identify and understand an avatar's intended reaction behaviors. They are discrete in their meaning for the reactions, removing the ambiguity that is present in physical gesturing and improving identification confidence as a result. Regardless of their effectiveness, however, EMOJI visualizations are not always preferable, as they remove the familiar, personal element of conversation gestures provide.

5.4 Combination of Gesture and Abstract Visuals

When comparing the combination of abstract visuals with gestures to their standalone equivalents, we again selected the more performant visualization condition. This was determined to be EXAGGERATED after analysis in Section 4. Thus the following discussion compares the EXAGGERATED, EMOJI, and EXAGGERATED & EMOJI visualizations.

We hypothesized that the combination of gestures and abstract visuals would be identified faster by participants than their standalone equivalents (*H3.1*). From recorded participant reaction report times and trial gaze summary data, our results do not support *H3.1*. While the combination of gestures and abstract visuals did significantly reduce the report times of participants for the Agree and Disagree reactions compared to the EXAGGERATED visualization, the same reactions were higher in report times than the EMOJI visualization, albeit without significance. Similarly in the gaze data, participants required significantly less time to stare at the avatar while reading with the combination visualization compared to the EXAGGERATED visualization but had significantly higher gaze percentage than the EMOJI visualization. We attribute this to the amount of visual information being conveyed to the participant by the avatar distracting them from the task. Participants reported this in post-study interviews, with P17 saying the combination visualization "kind of distracts" them while reading and P9 explicitly saying that the combination "is a little bit distracting from what I was reading there. It took me out of that (reading) a little bit to focus on the avatar."

We also hypothesized that the combination of gestures and abstract visuals would be identified with more confidence by participants than their standalone equivalents (*H3.2*). Based on participant identification confidence values, our results do not support *H3.2*. The reported identification confidence by participants had no significant differences between EMOJI, EXAGGERATED, and EXAGGERATED & EMOJI visualizations conveyed by the avatar across reactions. However, the combination of gestures had higher reported values across all reactions compared to the standalone EXAGGERATED and EMOJI conditions. We attribute this to the abstract visuals acting as a supplement to the physical gestures, aiding participants in distinguishing what reaction was conveyed by the avatar. In post-study interviews, participants reported this supporting benefit in the combination visualization. P7 described that the combination was "easy to understand and the emoji definitely helps with that too," P12 stated "it was easier to see the appearance of the emoji and also to see the avatar moving kind of draws the eye over," and P15 explained that the "emoji mixed with the exaggerated version of the reaction made it a lot easier to understand."

Finally, we hypothesized that the combination of gestures and abstract visuals would be preferred by participants over their standalone equivalents (*H3.3*). From participant visualization rankings, between-block questionnaire values, and post-study interview discussions, we were unable to determine whether *H3.3* was supported or not. Visualization rankings conducted in the post-study had 43% of participants identify the combination condition as their preferred visualization compared to both EXAGGERATED and EMOJI. Participants rated the combination visualization as significantly less natural than the EXAGGERATED visualization, but significantly more natural than the EMOJI visualization, which suggests that the gesture component benefits naturalness but the emoji component makes the combination less natural than the gesture alone. In addition, the combination was considered significantly more noticeable than the EXAGGERATED visualization, which suggests that the benefits of the emoji carry over to the combination condition. Participant post-study interviews suggest support of a combination of gestures with emojis. P1 stated that "the reaction, as well as the incorporation of an emoji, makes it very clear in my opinion," P2 expressed that understanding the reactions was "most effective with (exaggerated) movement plus emoji," and P26 described that "looking at the avatar first and seeing their reaction and then amplifying it with the sign made it more effective." More research regarding the combination of gestures and abstract visuals is needed to reject or support the hypothesis.

Overall, the combination of gestures with emojis had a mixed impact on the identification, understanding, and perception of participants compared to the standalone EXAGGERATED and EMOJI visualizations. While the combination did not necessarily outperform EXAGGERATED and EMOJI visualizations, it did provide the benefit of maintaining the naturalism of gestures and the clear identification of the intended meaning provided by emojis.

6 LIMITATIONS & FUTURE WORK

Our work had several limitations. First, the animations selected for the NATURAL and EXAGGERATED visual reactions were not consistent between the female and male avatars. This, along with the gender imbalance among the recruited participants and the findings from the training task, prompted us to exclude the gender variable from our results analysis. Future studies should ensure the balancing of participant gender and uniform animations by using a common auto-rigging tool when selecting animations. Second, the selection of NATURAL and EXAGGERATED visual reactions was subjective. The differences between the conditions relied on investigator judgment, which primarily was based on additionally perceived body movement (i.e., exaggerated Agree and Disagree included hand movements and the exaggerated question included lower body movement). Lastly, the sample size of our study was relatively small in comparison to most psychology studies [11, 31, 38]. A larger sample size may have enabled us to find significance in cases where values were trending towards a significant difference. Future studies analyzing the potential use of abstract visuals and comparisons of full-body avatar gesturing should include more participants.

Future directions for this work include the comparison of additional non-verbal cues, such as those relating to emotions [9] or a variety of affective states [29]. The reactions selected are a small subset of the possible affective displays people are capable of conveying in co-located conversation. Additionally, more visualizations and combinations should be considered, including the use of colors or pictorial runes to encode reaction intents. Lastly, further study should be conducted between two or more human participants working on a collaborative task. While we mimicked the distraction of shared material diverting attention between the participant and the avatar, this does not entirely capture the potential influence of the system for unscripted collaboration and communication.

7 CONCLUSION

We studied the effects of different non-verbal cues in the form of visual reactions on participants' identification, understanding, and perception of their intended meaning using avatars in VR. We found that exaggerated gestures present benefits in collaborative activities where shared materials may capture an observer's attention compared to equivalent non-exaggerated gestures. We also found that abstract visuals are an effective means to represent intentional reactions in regard to participant identification and understanding. Last, we discovered that the combination of gestures and abstract visuals, while not outperforming their standalone counterparts, provides comparable performance and understanding while preserving the naturalness of communication in a collaborative scenario. Our findings suggest further studies should be performed regarding the possibility of enhancing non-verbal cues using abstract visuals in collaboration and communicative scenarios.

ACKNOWLEDGMENTS

This work was supported by a grant from the Office of Naval Research.

REFERENCES

[1] L. Bartram, A. Patra, and M. Stone. Affective color in visualization. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, p. 1364–1374. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3025453.3026041 2

[2] M. Billinghurst and H. Kato. Collaborative augmented reality. *Communications of the ACM*, 45(7):64–70, 2002. 2

[3] R. R. A. Bourne et al. Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. *The Lancet Global Health*, 5(9):e888–e897, 2017. doi: 10.1016/S2214-109X(17)30293-0 4

[4] Z. Choudhary, A. Erickson, N. Norouzi, K. Kim, G. Bruder, and G. Welch. Virtual big heads in extended reality: Estimation of ideal head scales and perceptual thresholds for comfort and facial cues. *ACM Trans. Appl. Percept.*, 20(1), jan 2023. doi: 10.1145/3571074 2, 3

[5] S. Şemsioğlu, P. Karaturhan, and A. E. Yantaç. Emote: An interactive online tool for designing real-time emotional ar visualizations. In *13th Augmented Human International Conference*, AH2022. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3532525.3532527 2

[6] I. Damian, C. S. S. Tan, T. Baur, J. Schöning, K. Luyten, and E. André. [poster] exploring social augmentation concepts for public speaking using peripheral feedback and real-time behavior analysis. In *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 261–262, 2014. doi: 10.1109/ISMAR.2014.6948440 1

[7] T. Drey, P. Albus, S. der Kinderen, M. Milo, T. Segschneider, L. Chanzab, M. Rietzler, T. Seufert, and E. Rukzio. Towards collaborative learning in virtual reality: A comparison of co-located symmetric and asymmetric pair-learning. In *Proceedings of the 2022 CHI Conference on Human Factors*

in Computing Systems, CHI '22. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3491102.3517641 2
[8] B. Eerden. Anger in asterix: The metaphorical representation of anger in comics and animated films. In *Multimodal metaphor*, pp. 243–264. De Gruyter Mouton, 2009. 2
[9] P. Ekman. An argument for basic emotions. *Cognition and Emotion*, 6(3-4):169–200, 1992. doi: 10.1080/02699939208411068 9
[10] L. A. Elkin, M. Kay, J. J. Higgins, and J. O. Wobbrock. An aligned rank transform procedure for multifactor contrast tests. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, UIST '21, p. 754–768. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3472749.3474784 6
[11] B. Fischer and C. Herbert. Emoji as affective symbols: Affective judgments of emoji, emoticons, and human faces varying in emotional content. *Frontiers in Psychology*, 12, 2021. doi: 10.3389/fpsyg.2021.645173 2, 3, 9
[12] C. J. Forceville and E. Urios-Aparisi, eds. *Multimodal Metaphor*. De Gruyter Mouton, Berlin, New York, 2009. doi: doi:10.1515/9783110215366 2
[13] D. Gelb, A. Subramanian, and K.-H. Tan. Augmented reality for immersive remote collaboration. In *2011 IEEE Workshop on Person-Oriented Vision*, pp. 1–6, 2011. doi: 10.1109/POV.2011.5712368 1
[14] M. Gonzalez-Franco, E. Ofek, Y. Pan, A. Antley, A. Steed, B. Spanlang, A. Maselli, D. Banakou, N. Pelechano, S. Orts-Escolano, V. Orvalho, L. Trutoiu, M. Wojcik, M. V. Sanchez-Vives, J. Bailenson, M. Slater, and J. Lanier. The rocketbox library and the utility of freely available rigged avatars. *Frontiers in Virtual Reality*, 1, 2020. doi: 10.3389/frvir.2020.561558 4
[15] J. G. Grandi, H. G. Debarba, and A. Maciel. Characterizing asymmetric collaborative interactions in virtual and augmented realities. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 127–135, 2019. doi: 10.1109/VR.2019.8798080 2
[16] M. Hassenzahl, S. Heidecker, K. Eckoldt, S. Diefenbach, and U. Hillmann. All you need is love: Current strategies of mediating intimate relationships through technology. *ACM Trans. Comput.-Hum. Interact.*, 19(4), dec 2012. doi: 10.1145/2395131.2395137 1
[17] A. Irlitti, S. Von Itzstein, L. Alem, and B. Thomas. Tangible interaction techniques to support asynchronous collaboration. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 1–6, 2013. doi: 10.1109/ISMAR.2013.6671840 2
[18] N. Kaya and H. H. Epps. Relationship between color and emotion: a study of college students. *College Student Journal*, 38:396+, Sep 2004. 3. 2
[19] L. K. Kaye, S. Rodriguez-Cuadrado, S. A. Malone, H. J. Wall, E. Gaunt, A. L. Mulvey, and C. Graham. How emotional are emoji?: Exploring the effect of emotional valence on the processing of emoji stimuli. *Computers in Human Behavior*, 116:106648, 2021. doi: 10.1016/j.chb.2020.106648 3
[20] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The International Journal of Aviation Psychology*, 3(3):203–220, 1993. 4
[21] K. Kiyokawa, M. Billinghurst, S. Hayes, A. Gupta, Y. Sannohe, and H. Kato. Communication behaviors of co-located users in collaborative ar interfaces. In *Proceedings. International Symposium on Mixed and Augmented Reality*, pp. 139–148, 2002. doi: 10.1109/ISMAR.2002.1115083 2
[22] M. E. Latoschik, D. Roth, D. Gall, J. Achenbach, T. Waltemate, and M. Botsch. The effect of avatar realism in immersive social virtual realities. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, VRST '17. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3139131.3139156 2
[23] S. Y. Oh, J. Bailenson, N. Krämer, and B. Li. Let the avatar brighten your smile: Effects of enhancing facial expressions in virtual environments. *PloS one*, 11(9):e0161794, 2016. 2, 3
[24] M. Pakanen, P. Alavesa, N. van Berkel, T. Koskela, and T. Ojala. "nice to see you virtually": Thoughtful design and evaluation of virtual avatar of the other user in ar and vr based telexistence systems. *Entertainment Computing*, 40:100457, 2022. doi: 10.1016/j.entcom.2021.100457 2
[25] C. Pidel and P. Ackermann. *Collaboration in Virtual and Augmented Reality: A Systematic Overview*, pp. 141–156. 08 2020. doi: 10.1007/978-3-030-58465-8_10 1, 2
[26] T. Piumsomboon, A. Dey, B. Ens, G. Lee, and M. Billinghurst. The effects of sharing awareness cues in collaborative mixed reality. *Frontiers in Robotics and AI*, 6, 2019. doi: 10.3389/frobt.2019.00005 2
[27] T. Piumsomboon, G. A. Lee, J. D. Hart, B. Ens, R. W. Lindeman, B. H. Thomas, and M. Billinghurst. Mini-me: An adaptive avatar for mixed reality remote collaboration. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, p. 1–13. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3173574.3173620 2
[28] T. Piumsomboon, Y. Lee, G. A. Lee, A. Dey, and M. Billinghurst. Empathic mixed reality: Sharing what you feel and interacting with what you see. In *2017 International Symposium on Ubiquitous Virtual Reality (ISUVR)*, pp. 38–41, 2017. doi: 10.1109/ISUVR.2017.20 2
[29] J. A. Russell. A circumplex model of affect. *Journal of personality and social psychology*, 39(6):1161, 1980. 9
[30] A. Schäfer, G. Reis, and D. Stricker. A survey on synchronous augmented, virtual and mixed reality remote collaboration systems. *ACM Comput. Surv.*, apr 2022. Just Accepted. doi: 10.1145/3533376 2
[31] E.-M. Seidel, U. Habel, M. Kirschner, R. Gur, and B. Derntl. The impact of facial emotional expressions on behavioral tendencies in women and men. *Journal of experimental psychology. Human perception and performance*, 36:500–7, 04 2010. doi: 10.1037/a0018169 1, 9
[32] S. Şemsioğlu and A. E. Yantaç. Co-exploring the design space of emotional ar visualizations. In *HCI International 2021-Late Breaking Posters: 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings, Part I*, pp. 377–384. Springer, 2021. 2
[33] K. Shinohara and Y. Matsunaka. Pictorial metaphors of emotion in japanese comics. In *Multimodal metaphor*, pp. 265–296. De Gruyter Mouton, 2009. 2
[34] A. Valente, D. S. Lopes, N. Nunes, and A. Esteves. Empathic aurea: Exploring the effects of an augmented reality cue for emotional sharing across three face-to-face tasks. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 158–166, 2022. doi: 10.1109/VR51125.2022.00034 2
[35] M. E. Walker, D. Szafir, and I. Rae. The influence of size in augmented reality telepresence avatars. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 538–546, 2019. doi: 10.1109/VR.2019.8798152 2
[36] T. Waltemate, D. Gall, D. Roth, M. Botsch, and M. E. Latoschik. The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. *IEEE Transactions on Visualization and Computer Graphics*, 24(4):1643–1652, 2018. doi: 10.1109/TVCG.2018.2794629 2
[37] T.-Y. Wang, Y. Sato, M. Otsuki, H. Kuzuoka, and Y. Suzuki. Effect of full body avatar in augmented reality remote collaboration. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 1221–1222, 2019. doi: 10.1109/VR.2019.8798044 2
[38] S. Wiseman and S. J. J. Gould. Repurposing emoji for personalised communication: Why means "i love you". In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, p. 1–10. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3173574.3173726 2, 9
[39] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, p. 143–146. Association for Computing Machinery, New York, NY, USA, 2011. doi: 10.1145/1978942.1978963 6
[40] B. Yoon, H.-i. Kim, G. A. Lee, M. Billinghurst, and W. Woo. The effect of avatar appearance on social presence in an augmented reality remote collaboration. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 547–556, 2019. doi: 10.1109/VR.2019.8797719 2
[41] L. Ştefan. Immersive collaborative environments for teaching and learning traditional design. *Procedia - Social and Behavioral Sciences*, 51:1056–1060, 12 2012. doi: 10.1016/j.sbspro.2012.08.287 2