Exploring the Social Influence of Virtual Humans Unintentionally Conveying Conflicting Emotions

Zubin Choudhary* Nahal Norouzi† Austin Erickson‡ Ryan Schubert§ Gerd Bruder¶
Gregory F. Welch‖

University of Central Florida

ABSTRACT

The expression of human emotion is integral to social interaction, and in virtual reality it is increasingly common to develop virtual avatars that attempt to convey emotions by mimicking these visual and aural cues, i.e. the facial and vocal expressions. However, errors in (or the absence of) facial tracking can result in the rendering of incorrect facial expressions on these virtual avatars. For example, a virtual avatar may speak with a happy or unhappy vocal inflection while their facial expression remains otherwise neutral. In circumstances where there is conflict between the avatar's facial and vocal expressions, it is possible that users will incorrectly interpret the avatar's emotion, which may have unintended consequences in terms of social influence or in terms of the outcome of the interaction.

In this paper, we present a human-subjects study ($N = 22$) aimed at understanding the impact of conflicting facial and vocal emotional expressions. Specifically we explored three levels of emotional valence (unhappy, neutral, and happy) expressed in both visual (facial) and aural (vocal) forms. We also investigate three levels of head scales (down-scaled, accurate, and up-scaled) to evaluate whether head scale affects user interpretation of the conveyed emotion. We find significant effects of different multimodal expressions on happiness and trust perception, while no significant effect was observed for head scales. Evidence from our results suggest that facial expressions have a stronger impact than vocal expressions. Additionally, as the difference between the two expressions increase, the less predictable the multimodal expression becomes. For example, for the happy-looking and happy-sounding multimodal expression, we expect and see high happiness rating and high trust, however if one of the two expressions change, this mismatch makes the expression less predictable. We discuss the relationships, implications, and guidelines for social applications that aim to leverage multimodal social cues.

Index Terms: Human-centered computing—Human computer interaction (HCI)—HCI design and evaluation methods—User studies; Human-centered computing—Human computer interaction (HCI)—Interaction paradigms—Virtual reality

1 INTRODUCTION

Emotion perception, which plays a major role in social interactions [21, 31], can be defined as one's ability to "accurately recognize and appraise the emotional expressions and reactions of others" [21]. Emotions are typically conveyed via facial, vocal, and bodily cues [31, 33, 49, 60]. While facial expressions have generally been more researched [31, 33, 49, 60], multimodal expressions of emotion, such as the combination of facial and vocal emotions, can lead to a more accurate assessment of one's emotion [30].

In Virtual Reality (VR) avatars and agents [18] are usually intended to convey emotions via the virtual human's facial expressions and speech. The perceived facial expression, e.g., whether smiling, frowning, or neutral, can directly affect perceived emotions, which can in turn induce corresponding changes in the emotional state of the observer [36], and ultimately the outcome of an interaction. In particular, displays of positive emotions often result in positive effects [44, 47, 58]. For instance, interactions with smiling social partners lead to higher trust and more cooperation compared to non-smiling ones [47]; while displays of anger were linked to lower trustworthiness and lower chance of engagement in affiliative behaviors [13, 38]. Additionally, the intensity of such positive emotions can amplify the positive outcomes [52, 59]. For instance, a broad smile exhibited by a waiter as opposed to a minimal smile resulted in a larger amount of tips [52]. In another case, the intensity of a smile (broad vs. narrow smile) influenced participants' perception of warmth and competence. Humans with broad smiles were judged to be warmer but less competent than those with narrow smiles [59].

Misperceptions of the intended emotions of virtual humans can have real consequences, in particular in terms of *social influence*, which is defined as "a change in an individual's thoughts, feelings, attitudes, or behaviors that results from interaction with another individual or a group" [45]. In contrast to persuasion, which is considered an intentional behavior, social influence may be inadvertent or accidental [19], while it may subtly influence one's sense of trust and decision making [8]. An inherent point with respect to accidental influence is that we are usually unaware of the misperception—we tend to believe what we perceive without questioning it. Indeed, awareness of a problem would likely result in *conscious* uncertainty or even confusion, and conscious consideration of the intentions.

Despite tremendous advances in VR technologies, there are ample opportunities for misperceptions of emotions, despite the intentions. For example, relatively few VR systems include facial tracking, so no matter what facial expression is rendered it is likely wrong. Sometimes the magnitude or the intensity of the recognized expressions may not be accurate. Even with facial tracking, errors in the tracking can result in the rendering the wrong expression. In addition, display pixel density and contrast will impact the effective resolution of a display, which can impact the perceptions of the rendered expressions. Given the tremendous and rapidly growing number of head-mounted displays (HMDs) in consumer hands (tens of millions [2, 26]), and the increasingly widespread use of social VR platforms, the potential impact of misperceptions of intended emotions is significant and growing.

While there are ample opportunities for unconscious misperceptions based on visual perception (e.g., face tracking and display issues) there are relatively few risks of misperceptions of vocal expressions. Most modern VR system support relatively high-quality audio, both in terms of the quality of the speech and the fidelity of the audio system. As such we were particularly interested in the

*e-mail: zubinchoudhary@knights.ucf.edu
†e-mail: lili.n.nrz@gmail.com
‡e-mail: ericksona@knights.ucf.edu
§e-mail: rschuber@ist.ucf.edu
¶e-mail: bruder@ucf.edu
‖e-mail: welch@ucf.edu

effects of *mismatches* between facial expressions (visual indications of emotion) and vocal expressions (aural indications of emotion). In addition, because researchers have previously used head scaling techniques in part to mitigate misperceptions of facial perceptions [10], we were interested to see whether such head scaling could positively or negatively affect the outcomes of an interaction in the presence of mismatched facial and vocal indicators of emotion. These interests led us to the following two research questions:

- **RQ1**: What are the impacts of mismatched facial and vocal expressions on social influence?
- **RQ2**: Can head scale amplify/attenuate the effects of the facial expressions on social influence?

To explore these questions we undertook a human-subjects study ($N = 22$) where we examined the effects of three levels of happiness (unhappy, neutral, and happy) expressed in both visual (facial) and aural (vocal) forms, and three levels of head scales (down-scaled, accurate, and up-scaled). We show how different multimodal expressions have an effect on happiness and trust perception, while different head scales showed no effect. We discuss the relationships, implications and guidelines for social applications that aim to leverage head scaling and social cues.

The remainder of this paper is structured as follows: Section 2 presents previous work on emotion perception, virtual humans, and head scale. Section 3 describes our human-subject study in which we investigate our research questions. Section 4 presents our experimental results. Section 5 discusses these results in the context of previous work. Section 6 concludes the paper.

2 Related Work

Here we present previous work on emotion perception and the impact of mismatched expressions during interactions with real and virtual humans. We also present recent findings on scale manipulations of virtual humans motivated by the opportunity to enhance emotion perception in virtual environments.

2.1 Emotion Perception and its Influence on Trust and Decision Making

Emotions and expressions are crucial in social interactions, and as humans we use different modalities to signal and perceive them, including facial cues, vocal cues, or bodily cues. They are used either as an evaluation of an external or internal event [48] or used to communicate this evaluation to others (e.g., an infant crying for attention) [34]. Conversely, it can be used by others to draw situational information [24] and to inform decision-making [28]. For example, in a negotiation game, Antos et al. [3] found that participants selected partners with expressions which were congruent to their behavior. These partners were perceived more trustworthy than others. Another experiment by Khooshabeh et al. [29] conducted a similar negotiation game, in which they showed that participants felt more threatened when their partner displayed angry expressions. These experiments provide evidence for the *"Emotions As Social Information" (EASI)* model [57], which states that in ambiguous situations, we use emotions to make sense of them and their effect depends on the context in which the interaction takes place. Thus displaying positive emotion could elicit more cooperative behaviors while negative emotions could hinder and elicit competitive behaviors.

Focusing on virtual agents, previous research has found that humans are capable of accurately recognizing facial expressions of virtual agents similar to real humans [14, 25, 27]. Beyond the recognition of emotional expressions of virtual agents, many researchers have focused on the influence of such expressions on human participants' perceptions. In some cases, research has shown that the expression of positive emotion can lead to higher trust [16, 55]. For instance, in an investment game, an agent with a smiling voice was perceived as more trusting than a neutral voice—an effect that persisted even after the agent exhibited behaviors that indicated its lack of trustworthiness [55]. However, other research findings indicate that factors, such as interaction context and congruence of multimodal emotions can also influence participants' perceptions [11, 12, 54]. For instance, in the context of an iterated prisoner's dilemma task, positive expressions of an agent naturally led to more cooperation when the agent adopted a cooperative strategy and not when the positive emotion was used as part of the agent's competitive strategy [11].

2.2 Impact of Mismatched Expressions

In another example, compared to agents with positive facial and/or vocal expressions, the agent that exhibited neutral facial and vocal expressions led to higher trust in a desert survival task, which the authors suggested to be an influence of the study task [54].

Such mismatched emotions exhibited by different channels (e.g., happy voice and unhappy face) have received some attention in the area of virtual agents to understand how humans integrate emotions from different channels [54]. For instance, Mower et al. [40, 41] analyzed the interaction between different combinations of human voice and facial cues. They mismatched the audio and facial cues, and participants rated in terms of valence, activation and dominance. While the facial cues affected human perception, they found a strong audio bias. Their results indicate that people integrate natural audio cues and synthetic video cues only when the expression is matched. It has also been suggested that the audio and video channels express different components of emotions; specifically video channel expresses emotional valence (positive - negative) and the audio channel the emotional activation (high excitation - low excitation) [17, 22, 41].

However, such mismatches can unintentionally happen during interactions with virtual avatar interlocutors, such as interactions in social VR experiences. One of the reasons for the occurrence of such mismatches in social VR platforms is technology and design limitations [35, 51]. While current social VR platforms take advantage of the full-body tracking afforded by several VR devices, there is still a considerable gap in the control and communication of certain nonverbal behaviors, such as facial expressions, pose, posture, and more accurate finger and hand tracking [35, 51].

In the presence of conflicting social cues, it is possible that the user's interpretation of the person or avatar's emotion will be biased towards particularly strong cues (i.e., cues that are easier to identify). If this is the case, then strengthening or weakening a cue may cause the observer to interpret the persons' or avatar's emotion differently. Such techniques could potentially be used to help the user interpret the emotion correctly or in a particular manner. For instance, Tidd and Lockard demonstrated that participants' behavior towards their server in a dining context changed depending on the intensity of the smile exhibited by the server, where exhibiting a broader smile resulted in participants tipping more compared to conditions in which the server exhibited a slight smile [52].

For virtual contexts, there are less limitations on how cues can potentially be strengthened or weakened. For example, the big head technique was investigated by Choudhary et al [10], where the head scale of virtual avatars was changed to improve facial expression recognition when the avatar was at different viewing distances. At distances beyond 3 meters, participants were unable to discern the avatar's facial expressions, so they chose to up scale the head of the avatar relative to its natural size in order to remain accurate in identifying its expressions. Choudhary et al. [9] also investigated the effects of the Big Head technique on distance perception in VR, which revealed a significant effect of big heads on distance judgments only if heads were presented as floating objects in VR, but not when spatially anchored and attached to a human body at true scale. For this reason, the potential benefits of increased ease of

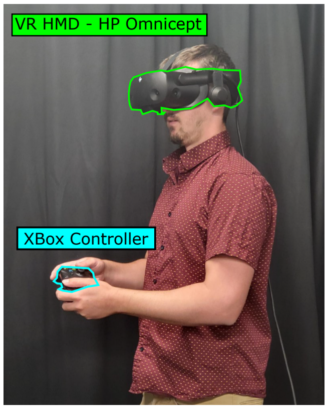


Figure 1: Participant with the HP Reverb G2 Omnicept Edition VR HMD and a wireless Xbox Controller.

recognizing facial expressions should be weighed against potential consequences of offset depth judgements, such as violating social proxemics norms [23]. An underlying assumption of the big head technique is that humans have a certain tolerance for seeing body parts at different sizes, and up-scaling certain parts can improve their respective effectiveness in conveying social cues. Aspects of this technique originated among game developers, who were looking for a solution to the problem of limited screen space and low pixel resolutions when trying to present game characters to players. Not surprisingly, some collaborative environments, such as Meta's Social VR application Horizons [42], have already started to leverage the approach, although the reasoning behind their design choices remains uncertain.

3 Experiment

To elucidate the effects of mismatched facial and vocal expressions on social influence, we designed an experimental scenario in VR around a virtual human that would recommend one of two options (A or B) to participants, who could only select one of them, not knowing if they could trust the virtual human's recommendation (see Figure 3). Similarly, as mismatches can originate in practical VR applications for different reasons, we designed the experimental stimuli so that participants may or may not notice a mismatch and if they noticed a mismatch they would not be certain about how to interpret it. The described experimental design and procedure was submitted to and approved by the institutional review board (IRB) of our university.

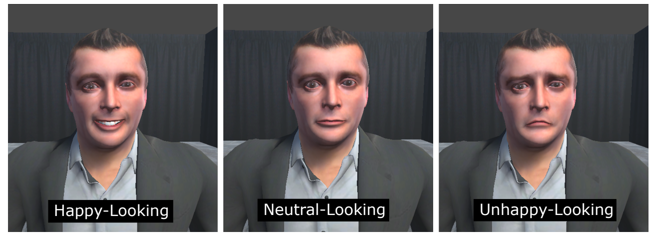
3.1 Participants

We recruited 22 participants from our university community: 15 male and 7 female; ages between 19 and 34, $M = 23.95$, $SD = 4.21$. All of the participants had normal or corrected-to-normal vision, 9 wore glasses and 1 wore contact lenses during the experiment. None of the participants reported any visual or vestibular disorders, such as color or night blindness, dyschromatopsia, or a displacement of balance. 20 participants had used a VR HMD before, and 13 of them had prior experience with social VR. The participants were either students or non-student members of our university community who responded to open calls for participation, and received monetary compensation for their participation. The experiment took participants on average 40 minutes to complete.

3.2 Material

3.2.1 Apparatus

Participants wore an immersive VR HMD, the HP Reverb G2 Omnicept Edition. The HMD provided a 90 degree vertical and 98 degree



(a) Facial Expressions

(b) Head Scales

Figure 2: Visual stimuli in our experiment.

horizontal field of view, and had a resolution of 2160×2160 pixels per eye at a refresh rate of 90 Hz. The HMD was connected to a host PC (Intel Core i9-10850K CPU @ 3.60 GHz, 64Gb Ram, NVIDIA RTX 3090 graphics card, Windows 10 Enterprise) that the experimenter used to run the Unity application (version 2020.3.2f1 LTS) and monitor the participant's view and activities within the virtual environment. Participants were instructed to stand on a marked location in the center of our lab space, clear of obstacles. Participants used a wireless Xbox controller (see Figure 1) to indicate responses to the visual stimuli in the experiment.

3.2.2 Virtual Stimuli

We adopted a life-size 3D male virtual human model from the Microsoft RocketBox Avatar Library [20] to use as the virtual human for our study (see Figure 3). A similar male model was used by Choudhary et al. [9], where they investigated the effects of head scale on distance estimation. Following their method, we leveraged custom blendshapes, for the three facial expressions (unhappy, neutral and happy), in Blender [6] to control the virtual human's head scale. The model was imported into the Unity game engine and was positioned in a virtual hallway environment to be 5 meters away from the participant. The virtual hallway had dimensions of 5 m (width) × 3 m (height) × 10 m (length).

For the box selection task, we positioned two virtual cubes between the virtual human and the participant; the cubes were labeled "A" and "B" and had the dimensions 20 cm (width) × 20 cm (height) × 20 cm (length). The cubes were 3 meters away from the participant and 1.5 meters horizontally apart from each other. We programmed the behavior and animation of the virtual human so that he could point at either cube and speak prerecorded dialogue to recommend that participants choose one of the two cubes. We used Rogo Digital Lip Sync Pro [46] to generate lip movements for the virtual human that matched his spoken dialogue. We used the Meta Quest Viseme References [37] to create custom blendshapes for lip sync. Throughout the experiment, the virtual human was made to blink at regular intervals and performed a looping idle animation from Adobe Mixamo [1].

The virtual human was capable of presenting facial expressions that appeared happy, neutral, and unhappy (see Figure 2(a)). To control the virtual human's facial expressions, we created custom blendshapes using Blender [15]. The virtual human was further capable of presenting vocal expressions that sounded happy, neutral, and unhappy. Therefore, we used prerecorded speech prompts by a male native English speaker who we trained to express the corresponding emotions. The speech prompts consisted of seven variations of the sentence *"For this condition, I choose this box!"* with minor differences in the sentence structure, each recorded for the three vocal expressions, which were accompanied by the character pointing at one of the cubes (A or B). To record the audio, we used a professional Blue Yeti USB microphone and equalized the volume of the recordings with the Audacity recording software [4]. In order to confirm that the facial expressions and recorded phrases conveyed the intended emotions, we had them independently reviewed and confirmed by three in-house participants. This was further confirmed through the analysis of the experiment data (i.e., see the happiness scores in Section 4.1).

3.2.3 Response User Interfaces

As shown in Figure 3, we included three user interfaces (UIs) to assess participants' responses, which were presented to participants as part of the virtual environment after the stimuli were presented. The UIs were fixed at two meters in front of the participant and positioned in the lower part of their visual field. The UIs included a label indicating the task, along with either a 7-point slider (trust scale or emotion scale) or two A/B buttons. Participants navigated the UIs with the joystick and buttons on the wireless Xbox controller.

3.3 Methods

We used a 3×3×3 full-factorial within-subjects design with three factors (*head scale*, *facial expression*, and *vocal expression*) as described below. We grouped the 27 conditions into 3 groups based on the 3 different head scales. These groups were presented to participants in a counterbalanced order through the use of a 3×3 Latin square. Within each group, the 3×3 expressions were presented in a randomized order to each participant.

We investigated the following independent variables:

- **Facial Expression (3 levels):** We chose three different facial expressions exhibited by the virtual human in the experiment: happy, neutral, unhappy (see Table 1).
- **Vocal Expression (3 levels):** We chose three different vocal expressions exhibited by the virtual human: happy, neutral, unhappy (see Table 1).
- **Head Scale (3 levels):**

  For each condition, the head scale was within one of three ranges:
  - **Accurate:** This range specifies common human head scales. Head scales were randomized to be within ±5% of our virtual human's default head scale.
  - **Down-scaled:** This range specifies head scales that are smaller than the default scale for the virtual human model. Head scales were randomized to be within 60–80% of the original head scale.
  - **Up-scaled:** This range specifies head scales that are larger than the default scale for the virtual human model. Head scales were randomized to be within 140–180% of the original head scale.

These ranges are based on the previous work by Choudhary et al. [10], which investigated comfortable ideal ranges of head scales for non-verbal communication (facial expressions). We did not use their exact reported ranges due to different experiment design and implementation, such as the type of VR HMD used and the representation of the virtual human. Their results provided evidence of an asymmetric range for perceived comfortable *down-scaled*, *accurate* and *up-scaled* heads. They found the widest ranges for *up-scaled* heads, followed by *down-scaled* heads, and smallest for *accurate* heads. Hence we opted for a similar asymmetric range in our experiment.

We introduced tags for the different combinations of facial and vocal expressions (see Table 1). The tags (e.g., U_fU_v) consist of two capital letters representing the conveyed emotion, U, N, or H, for unhappy, neutral, and happy, respectively. The subtext beneath each capital letter, $_f$ or $_v$, represent whether its the facial expression or vocal expression. We used the colors to highlight the unhappy (red), neutral (gray), and happy (green) emotions.

Table 1: This table presents the nine different multimodal emotional expressions exhibited by the virtual human in the experiment. Conditions on the diagonal have matching emotions in the facial expression and vocal inflection, whereas conditions off of the diagonal have mismatched combinations of emotions being presented.

	Facial Expression		
Vocal expression	**Unhappy**	**Neutral**	**Happy**
Unhappy	U_fU_v	N_fU_v	H_fU_v
Neutral	U_fN_v	N_fN_v	H_fN_v
Happy	U_fH_v	N_fH_v	H_fH_v

3.3.1 Procedure

Upon arrival, participants read through a description of the study procedure and consent form, and were asked to give their verbal consent to participate in the experiment.

The experimenter then described the experimental scenario to participants, explaining that it involves a virtual human who will recommend one of two options (A or B) to them, but that they can only select one of them, not knowing if they can trust the virtual human's recommendation. They would be asked to not only select either option A or B, going with or against the virtual human's recommendation, but also rate their impression of how trustworthy that virtual human is on a 7-point scale.

In an effort to not bias the participants to expect certain emotions from the virtual human, the experimenter did not explicitly describe the individual stimuli – unhappy, neutral and happy facial expressions and voices, and head scales (see Section 3.3).

Participants then donned the HP Omnicept HMD, and began the calibration, which took an average of 3 minutes to complete. The calibration guided the participants to adjust the interpupillary distance (IPD) and confirmed that the headset was positioned optimally for the best viewing experience. After calibration, the experimenter started the Unity application on the headset from the connected computer. Participants then saw the virtual human, facing them, standing behind the two floating cubes in the virtual hallway. Before experiencing the study conditions, participants completed a practice session in VR to familiarise themselves with the task and UI functionality with the Xbox controller, as described in Section 3.2.3. Participants took on average 3 minutes to become familiar and then the first of the experimental trials was started.

For each trial, participants observed the virtual human exhibiting a facial expression according to the current condition, who was standing behind cubes "A" and "B". This virtual human recommended that the participant choose between the two cubes by stating aloud a variation of the sentence "For this condition, I choose this box," which was spoken with the vocal expression according to the current condition. As the virtual human was saying this, he would point to one of the cubes, chosen at random. Following this, a UI appeared that asked the participant to rate the trustworthiness of the virtual human (see Section 3.4). The UI then asked the participant to choose between the two A/B cubes. Once the participant answered the two UI prompts, they moved on to the next condition. Conditions progressed in a similar manner until the participant had completed all 27 conditions.

After completing the 27 conditions, participants experienced the conditions in the same order a second time. During this second run-through, the stimuli were identical but the UI asked participants to rate the emotion perceived from the virtual human, instead of

Figure 3: Screenshots showing the response user interfaces we used in the experiment: trust scale, A/B selection task, and emotion scale.

asking about perceived trust and which cube they would select. This repetition of conditions was performed so that participants would not establish a relationship between happiness and trust, which could have affected their initial trust responses.

After completing all conditions and tasks, participants proceeded to complete a post-questionnaire, assessing their demographics, prior VR experience, and their general perception of the virtual human. In total the experiment took an hour to complete and the experiment ended with the participant receiving monetary compensation for their time.

3.4 Measures

In this section, we describe the four main measures in the experiment: the trust scores, box selections, happiness scores, and fixation time.

3.4.1 Trust Scores

In each trial, after the virtual human recommended cube "A" or "B" to the participant, we asked participants to rate their trust in the virtual human on a 7-point Likert scale (1 = Untrustworthy to 7 = Trustworthy) as an explicit measure of trust.

3.4.2 Box Selections

As an implicit measure of the participant's trust in the virtual human, we recorded their responses in the box selection task relative to whether or not the participant chose the cube that was recommended by the virtual human (0 = Recommendation not followed to 1 = Recommendation followed).

3.4.3 Happiness Scores

For each combination of facial expression and vocal expression, we asked participants to rate the virtual human's expressed happiness on a 7-point Likert scale (1 = Unhappy to 7 = Happy).

4 Results

We analyzed the responses with a 3 (head scales) $\times$ 3 (facial expressions) $\times$ 3 (vocal expressions) three-way repeated-measures ANOVAs and Tukey multiple comparisons with Bonferroni correction at the 5% significance level. We tested the assumptions of the parametric statistical tests. We confirmed the normality with Shapiro-Wilk tests at the 5% level and QQ plots. Degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity when Mauchly's test indicated that the assumption of sphericity had been violated.

4.1 Happiness Scores

The results for *Happiness Scores* are shown in Figures 4(a)-(c) and 5(a).

Facial Expressions: We found a significant main effect of facial expressions on *Happiness Scores*, $F(2,42) = 142.4$, $\boldsymbol{p < 0.001}$, $\eta_p^2 = 0.87$. Post-hoc tests showed that all pairs were significant, $U_f{<}N_f$, $U_f{<}H_f$, and $N_f{<}H_f$.

Vocal Expressions: We found a significant main effect of vocal expressions on *Happiness Scores*, $F(1.39,29) = 162.4$, $\boldsymbol{p < 0.001}$, $\eta_p^2 = 0.88$. Post-hoc tests showed that all pairs were significant, $U_v{<}N_v$, $U_v{<}H_v$, and $N_v{<}H_v$.

Head Scale: We did not find any significant effect of head scale on *Happiness Scores*, nor any interaction effect between head scale and expressions.

Interaction Between Facial and Vocal Expressions: We found a significant interaction effect between facial and vocal expressions on *Happiness Scores*, $F(4,84) = 4.74$, $\boldsymbol{p = 0.002}$, $\eta_p^2 = 0.19$. Pairwise comparisons across the two factors showed that all pairs of facial and vocal expressions were significant, except between the following pairs, U_fH_v and N_fU_v, U_fH_v and H_fU_v, N_fN_v and H_fU_v, and N_fH_v and H_fN_v. Overall, U_fU_v expression had the lowest happiness scores and H_fH_v had the highest happiness scores.

4.2 Trust Scores

The results for *Trust Scores* are shown in Figures 4(d)-(f) and 5(b).

Facial Expressions: We found a significant main effect of facial expressions on *Trust Scores*, $F(1.36,28.55) = 5.7$, $\boldsymbol{p = 0.016}$, $\eta_p^2 = 0.21$. Post-hoc tests showed the following two pairs significant, $U_f{<}N_f$, and $U_f{<}H_f$.

Vocal Expressions: We found a significant main effect of vocal expressions on *Trust Scores*, $F(1.55,32.51) = 17.67$, $\boldsymbol{p < 0.001}$, $\eta_p^2 = 0.46$. Post-hoc tests showed all pairs significant, $U_v{<}N_v$, $U_v{<}H_v$, and $N_v{<}H_v$.

Head Scale: We did not find any significant effect of head scale on *Trust Scores*, nor any interaction effect between head scale and expressions.

Interaction Between Facial and Vocal Expressions: We found a significant interaction effect between facial and vocal expressions on *Trust Scores*, $F(2.82,59.2) = 2.99$, $\boldsymbol{p = 0.041}$, $\eta_p^2 = 0.13$. Pairwise comparisons across the two factors showed significant differences only between the following pairs, $U_fU_v{<}N_fH_v$, $U_fU_v{<}H_fH_v$, $U_fN_v{<}N_fH_v$, $N_fU_v{<}N_fN_v$, $N_fU_v{<}N_fH_v$, $N_fU_v{<}H_fH_v$, $H_fU_v < N_fH_v$, and $H_fU_v{<}H_fH_v$. These findings and the overall trend in Figure 5(b) suggest an inclination to higher trust for neutral and happier expressions as opposed to unhappier or strongly mismatched expressions.

4.3 Box Selections

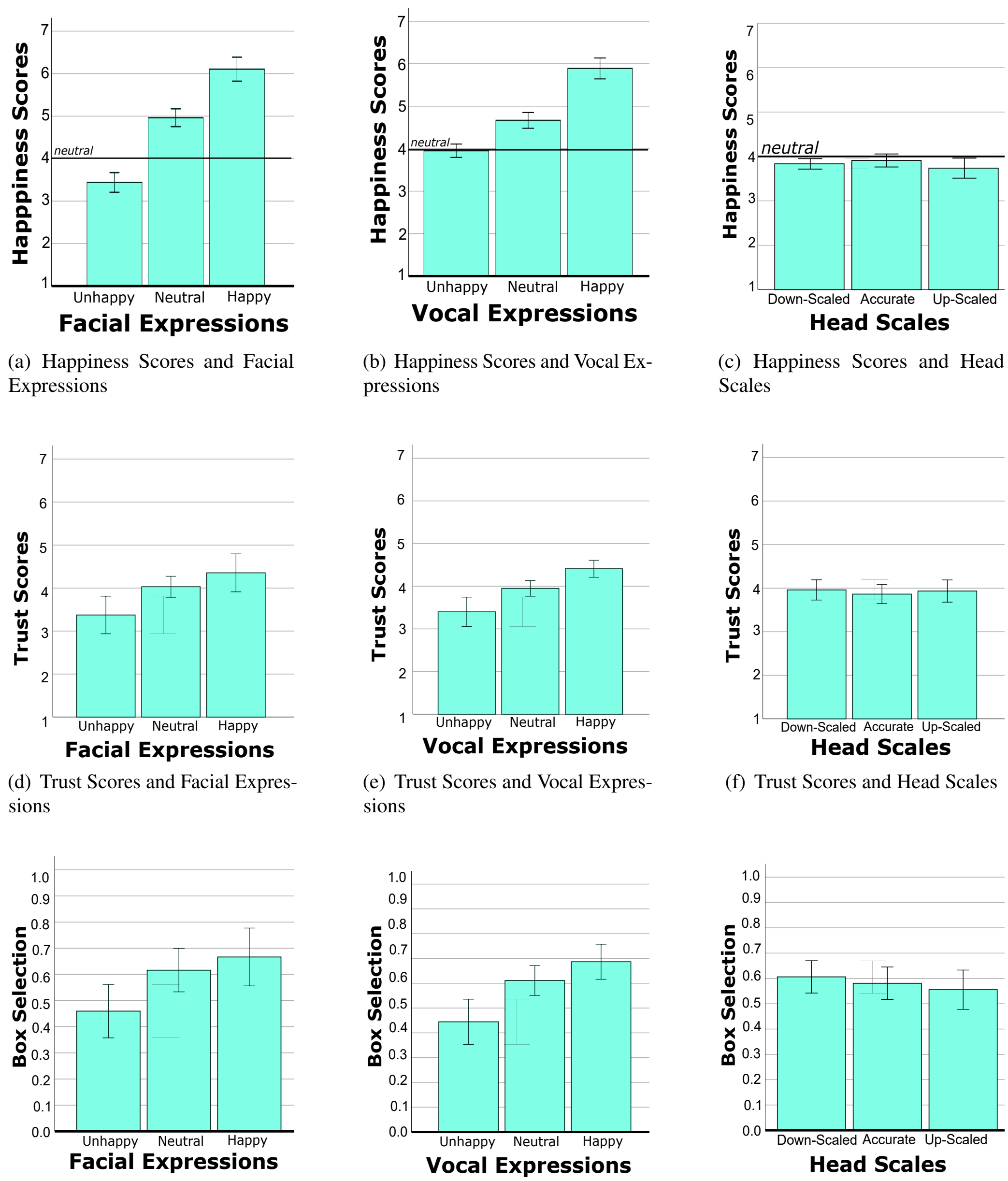
The results for *Box Selections* are shown in Figures 4(g)-(i) and 5(c).

Facial Expressions: We found a significant main effect of facial expressions on *Box Selections*, $F(1.48,31) = 4.25$, $\boldsymbol{p = 0.034}$, $\eta_p^2 = 0.17$. Post-hoc tests showed only one pair significant, $U_f{<}N_f$.

Vocal Expressions: We found a significant main effect of vocal expressions on *Box Selections*, $F(2,42) = 12$, $\boldsymbol{p < 0.001}$, $\eta_p^2 = 0.36$. Post-hoc tests showed two pairs significant, $U_v{<}N_v$ and $U_v{<}H_v$.

Head Scale: We did not find any significant effect of head scale on participants' *Box Selections*, nor any interaction effect between head scale and expressions.

Interaction Between Facial and Vocal Expressions: We do not see a significant interaction effect between facial and vocal expressions on *Box Selections*.



(a) Happiness Scores and Facial Expressions

(b) Happiness Scores and Vocal Expressions

(c) Happiness Scores and Head Scales

(d) Trust Scores and Facial Expressions

(e) Trust Scores and Vocal Expressions

(f) Trust Scores and Head Scales

(g) Box Selections and Facial Expressions

(h) Box Selections and Vocal Expressions

(i) Box Selections and Head Scales

Figure 4: Experimental results for our main effects. The x-axes show the three facial expressions, vocal expressions and head scales. The y-axes show the (a)-(c) happiness scores (1=Unhappy to 7=Happy), (d)-(f) trust scores (1=Untrustworthy to 7=Trustworthy), (g)-(i) participants' Box selections relative to the virtual human's recommendation (0=Recommendation not followed to 1=Recommendation followed).

4.4 Correlations

We ran Pearson correlations for the following pairs of two variables to assess the strength of the linear relationship between them.

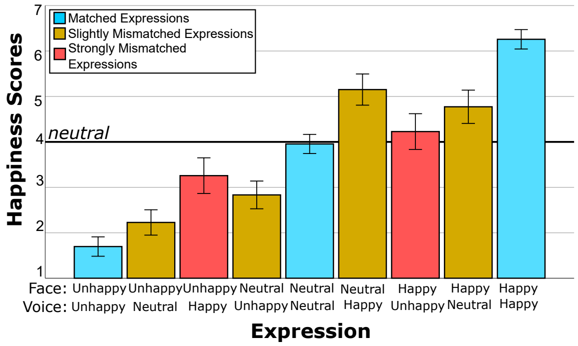
Between Happiness Scores and Trust Scores We found a positive correlation between the two variables, $r = 0.39$, $\boldsymbol{p < 0.001}$, indicating that happiness and trust results are correlated.

Between Trust Scores and Box Selections We found a positive correlation between the two variables, $r = 0.65$, $\boldsymbol{p < 0.001}$, indicating that trust and box selections are correlated.

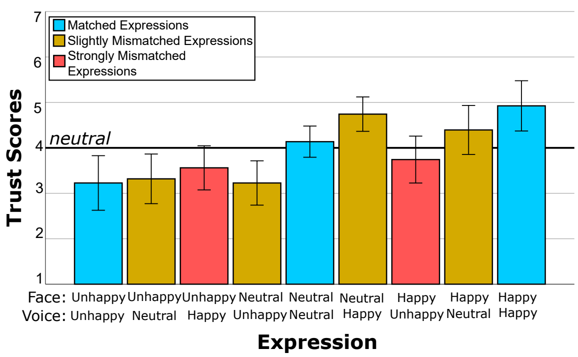
Between Happiness Scores and Box Selections We found a positive correlation between the two variables, $r = 0.19$, $\boldsymbol{p < 0.001}$, indicating that happiness and box selections are correlated.

5 Discussion

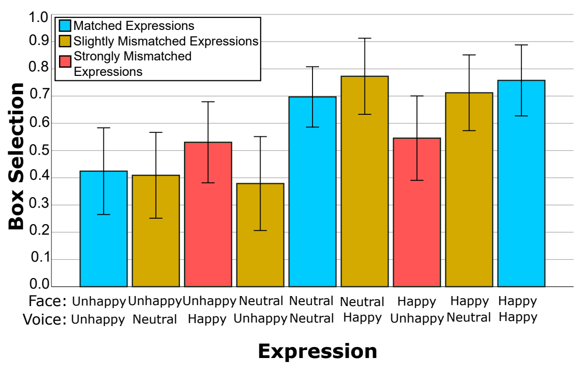
In this section, we summarize our main findings and discuss effects of mismatched facial and vocal expressions and head scales, while also presenting implications and addressing limitations of our experiment.



(a) Happiness Scores



(b) Trust Scores



(c) Box Selections

Figure 5: Experimental results for our interaction effects. The x-axes show our tags for the expressions of the virtual human (see Table 1). The y-axes show the (a) happiness scores (1=Unhappy to 7=Happy), (b) trust scores (1=Untrustworthy to 7=Trustworthy), and (c) participants' box selections relative to the virtual human's recommendation (0=Recommendation not followed to 1=Recommendation followed).

5.1 A Virtual Human's Mismatched Expressions Impact Perception and Social Influence

As discussed above, virtual humans come in different types and are used for different purposes, such as *avatars* where one's facial expression are often not tracked by one's HMD [35, 51], or *agents* where their behavioral model may or may not account for emotions when generating speech or facial expressions [43]. These may lead to virtual humans being able to present none or only either facial or vocal expressions, e.g., practitioners may set an avatar's face to a neutral (or happy) default expression if no face tracking is available. However, even if face tracking is available, in certain situations one's expression of emotion on one's face or in one's voice may differ, e.g., when one tries to convey an emotion one does not actually feel, which may be difficult to distinguish from technological limitations in VR. We discuss our results for happiness perception and trust in the following with a view on such matched or mismatched facial and vocal expressions of a virtual human.

Happiness Perception Among our tested conditions, we evaluated three multimodal expressions that agreed with each other (i.e., matching facial and vocal expressions). As expected, our results in Section 4.1 show that the matched unhappy expression $U_f U_v$ was rated the unhappiest among all conditions, our matched happy expression $H_f H_v$ was rated the happiest, and our matched neutral expression $N_f N_v$ was rated as neutral (happiness score close to 4 on our 1 to 7 scale; see Figure 5(a), with significant effects $U_f U_v < N_f N_v < H_f H_v$.

When both facial and vocal expressions were strongly mismatched (i.e., opposite expressions) for $U_f H_v$ and $H_f U_v$, our participants estimated them as significantly less happy/unhappy (i.e., more neutral) than with matching scores: $U_f U_v < U_f H_v$ and $H_f U_v < H_f H_v$. We found no significant differences between either of them and our condition with matched neutral expressions $N_f N_v$. Based on related work from Mower et al. [40], who found that for conflicting multimodal expressions the vocal channel may dominate the visuals, our initial assumption was that we may find a similar effect here, i.e., $H_f U_v < U_f H_v$, but our results do not support this assumption.

In the following pair with a slight mismatch, $U_f N_v < N_f U_v$, the facial and vocal expressions include each one unhappy and neutral expression, but the condition with an unhappy facial expression was estimated as more unhappy than the one with an unhappy vocal expression, suggesting that the facial cues in this case had a stronger effect than the vocal cues. However, for the pair, $H_f N_v$ and $N_f H_v$, neither a happy facial expression nor a happy vocal expression had a stronger effect when the other modality was neutral.

For slight mismatches, we overall found $U_f N_v < N_f U_v$, $N_f N_v < H_f N_v$ and $N_f H_v$, which is interesting because when either modality of the expression changed from neutral, the overall perception of the multimodal expression also changed significantly, except when the vocal expression changed from neutral to unhappy, $N_f U_v$ and $N_f N_v$. This suggests that the influence of the unhappy voice was not strong enough to elicit a significant difference from a neutral matched expression and it had a weaker effect than an unhappy facial expression.

As discussed in Section 1, the transmission of emotions by a virtual human may not be accurate. For example, when users embody avatars in a shared VR environment without an HMD capable of face-tracking (i.e., where the facial expression remains static even when users convey different emotions vocally). In these cases, we find that the choice of one's default facial expression of the avatar matters. If practitioners choose to set the default expression to a neutral facial expression, our results shows $N_f N_v < N_f H_v$ If they choose a happy facial expression, we found $H_f U_v < H_f N_v < H_f H_v$, and if they choose an unhappy facial expression, we also found $U_f U_v < U_f N_v < U_f H_v$. However, as expected, Figure 5(a) shows that choosing a happy (H_f) or unhappy (U_f) facial expression causes a bias in the overall estimation of happiness. Overall, compared to the effects for matching expressions $U_f U_v < N_f N_v < H_f H_v$, all of these results with a static facial expression indicate that the *dynamic range* of the estimated happiness scores of the virtual human is greatly reduced, emphasizing the importance of face-tracking for practitioners in VR if the goal is to support veridical emotion perception. This also suggests an overall higher impact on happiness perception by facial expressions over vocal expressions.

Trust Perception To understand the participants overall trust perception, we use results from their trust scores and relative A/B selections (refer Section 4.2 and 4.3). The trust scores were used as an explicit measure while the relative A/B selections were used as an implicit measure of trust. Section 4.4 showed a positive correlation between the two trust measures, which signifies a positive non-linear relationship ($r = 0.65$, $\boldsymbol{p < 0.001}$). Since we did not find a strong positive correlation, we found significant results and pairs from our trust scores that were not found from the box selections, such as the interaction effect between facial and vocal expressions. We believe this happened because of the simplicity of the task and does not represent the choices we might need to make in our everyday lives. Other studies measuring trust, such as Torre et al. [55, 56], used a version of the survival task called "Lunar survival task" while others used trust games [5]. Trust in itself is connected to multiple interpersonal phenomena – persuasion, social status, knowledgeableness and more, and such tasks are more representative of that than a binary box selection task.

From our results, among our three agreeing multimodal expressions, the unhappy expression U_fU_v was rated significantly less trustworthy than the happy expression H_fH_f.

We also found $N_fU_v < N_fN_v$, $N_fU_v < N_fH_v$ and $N_fU_v < H_fH_v$ on our trust scores. N_fU_v is an interesting condition because it resembles a situation where facial tracking is absent and the user sounds unhappy. Its counterpart condition N_fH_v, seems to have been perceived highly trustworthy. To support this, we also find $U_fN_v < N_fH_v$, $N_fU_v < N_fH_v$, and $H_fU_v < N_fH_v$ on trust scores. From Figure 5(b) and (c), we can see very low trust responses for N_fU_v and high trust responses for N_fH_v, which suggests that when the facial expression was neutral, voice alone is capable of making the virtual human both trustworthy (happy vocals) and untrustworthy (unhappy vocals). Similar experiments by Torre et al. [53, 54] found N_fN_v and H_fN_v to be the most trustworthy respectively. Our results neither support nor disprove their results, we speculate the gap may be because of the highly different contexts of the simulated scenarios and used technologies.

For practitioners using virtual humans, having matching multimodal expressions is more predictable. Happy virtual humans are perceived happier and more trustworthy, while the opposite for unhappy virtual humans. Other mismatched and conflicting expressions are less predictable in terms of what to expect from them. We discussed several significant pairs but without a common trend. So, virtual humans with matched expressions are more reliable than ones with mismatched or conflicting expressions.

5.2 Head Scale Does Not Significantly Affect Social Influence

Our results suggest that manipulated head scales did not have any effect on participants' estimation of the virtual human's happiness expressions or our implicit and explicit trust measures. In line with the *Uncanny Valley* theory [39], which proposes that increasing a human representation's visual realism may not necessarily result in an increased sense of comfort when interacting with that entity, potential Uncanny Valley effects could be caused by our virtual avatar with disproportionate heads. Additionally, the scaled or altered heads reduce the virtual human's humanness and may have given rise to a new sense of "creepiness" or "eeriness." Despite exaggerating the expressions and emotions on the face, the above explanations could have hindered our ability to see their effects on emotion and trust perception in our experiment. This is still interesting because even though participants perceived the expressions of virtual humans as expected, the expressions did not have a significant effect on participants' estimations of happiness and trust.

For practitioners in VR, these results suggests that head scale, e.g., the up-scaled heads used in Meta's Horizon Worlds or Workrooms, do not necessarily introduce a bias to emotion perception or social influence.

5.3 Limitations and Future Work

Our study provided evidence on the importance of facial and vocal expression agreement, specifically on happiness and trust. However, there are also a few limitations of the current work, which can lead us to interesting research vistas that may be investigated in the future.

In this study, we chose to use a stylized male virtual human in a static setting with idle standing animations. In social VR, avatars may have different rendering styles and appearances [32, 61] in a more dynamic environment. Prior work suggests that the render style affects the personality of virtual humans [62], and cartoon-like avatars have been shown to be more trustworthy than photorealistic avatars [54]. We also explore only one emotional state (happiness) and two modalities of emotional expression (facial and vocal expressions). Other cues that would be interesting to investigate are body movement, posture, gestures, and eye contact [7, 50]. We believe different styles, and mismatches of other expressive modalities are interesting research directions to explore in the context of emotion perception and social influence.

The box selection task to measure trust is more simplified compared to other popular trust tasks, such as negotiation tasks, trust games [5], survival tasks [56], etc. We chose this design as it allowed us to control and evaluate multiple conditions in VR in under one hour. For more rigorous testing, other trust tasks (not limited to the above mentioned) could be used to further understand trust perception with mismatched expressions and different head scales.

Our study sample further included more young male adults, and more representative samples in future work should provide a more comprehensive picture of how people perceive virtual humans with different genders with matched or mismatched emotions and up-scaled or down-scaled heads.

Last but not least, our study focused on assessing participants' impressions of the virtual human after a comparatively brief interaction. Interesting research questions arise for the influence of expressions and head scale manipulations in longitudinal studies.

6 Conclusion

In this paper, we presented a human-subjects study that investigated the importance of facial and vocal expression agreement and head scale on happiness perception and social influence. We found evidence that for matched facial and vocal expressions, our participants rated the matched happy condition as the happiest and the matched unhappy condition as the unhappiest, while for the slightly or strongly mismatched expressions, our participants rated them in between the range of matched expressions, effectively reducing their dynamic range. Further, our participants trusted matched happy expressions more than matched unhappy expressions, while we also found evidence that when the facial expressions were neutral, unhappy and happy audio expressions were capable of swaying participants' trust negatively and positively, respectively. We discussed implications of our results, limitations of our study design, and future work.

Acknowledgments

This material includes work supported in part by the National Science Foundation under Award Numbers 1564065 and 1800961 (Dr. Ephraim P. Glinert, IIS) to the University of Central Florida; the Office of Naval Research under Award Number N00014-18-1-2927 (Dr. Peter Squire, Code 34); and the Advent Health Endowed Chair in Healthcare Simulation (Prof. Welch).

References

[1] Adobe Mixamo. Animate 3D characters for games, film, and more. `https://www.mixamo.com/`.
[2] T. Alsop. Virtual reality (vr) headset unit sales worldwide from 2019 to 2024. `https://www.statista.com/statistics/677096/vr-headsets-worldwide/`, Jan 25 2022.

[3] D. Antos, C. De Melo, J. Gratch, and B. J. Grosz. The influence of emotion expression on perceptions of trustworthiness in negotiation. In *Twenty-Fifth AAAI Conference on Artificial Intelligence*, 2011.

[4] Audacity. Free, open source, cross-platform audio software. `https://www.audacityteam.org/`.

[5] J. Berg, J. Dickhaut, and K. McCabe. Trust, reciprocity, and social history. *Games and economic behavior*, 10(1):122–142, 1995.

[6] Blender. Free and open-source 3D computer graphics software. `https://www.blender.org/`.

[7] J. K. Burgoon, X. Wang, X. Chen, S. J. Pentland, and N. E. Dunbar. Nonverbal behaviors "speak" relational messages of dominance, trust, and composure. *Frontiers in psychology*, 12:7, 2021.

[8] T. R. Campellone and A. M. Kring. Who do you trust? the impact of facial emotion and behaviour on decision making. *Cognition & emotion*, 27(4):603–620, 2013.

[9] Z. Choudhary, M. Gottsacker, K. Kim, R. Schubert, J. Stefanucci, G. Bruder, and G. F. Welch. Revisiting distance perception with scaled embodied cues in social virtual reality. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 788–797. IEEE, 2021.

[10] Z. Choudhary, K. Kim, R. Schubert, G. Bruder, and G. F. Welch. Virtual big heads: Analysis of human perception and comfort of head scales in social virtual reality. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 425–433. IEEE, 2020.

[11] C. M. de Melo, J. Gratch, and P. J. Carnevale. Humans versus computers: Impact of emotion expressions on people's decision making. *IEEE Transactions on Affective Computing*, 6(2):127–136, 2014.

[12] C. M. de Melo and K. Terada. Cooperation with autonomous machines through culture and emotion. *PloS one*, 14(11):e0224758, 2019.

[13] J. R. Dunn and M. E. Schweitzer. Feeling and believing: the influence of emotion on trust. *Journal of personality and social psychology*, 88(5):736, 2005.

[14] M. Dyck, M. Winbeck, S. Leiberg, Y. Chen, R. C. Gur, and K. Mathiak. Recognition profile of emotions in natural and virtual faces. *PloS one*, 3(11):e3628, 2008.

[15] P. Ekman and W. V. Friesen. Facial action coding system. *Environmental Psychology & Nonverbal Behavior*, 1978.

[16] A. C. Elkins and D. C. Derrick. The sound of trust: voice as a measurement of trust during interactions with embodied conversational agents. *Group decision and negotiation*, 22(5):897–913, 2013.

[17] S. Fagel. Emotional mcgurk effect. In *Proceedings of the international conference on speech prosody*, vol. 1. Citeseer, 2006.

[18] J. Fox, S. J.-G. Ahn, J. Janssen, L. Yeykelis, K. Segovia, and J. Bailenson. Avatars Versus Agents: A Meta-Analysis Quantifying the Effect of Agency on Social Influence. *Human-Computer Interaction*, 30:401–432, 2015.

[19] R. H. Gass. Social influence, sociology of. In J. D. Wright, ed., *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)*, pp. 348–354. Elsevier, second edition ed., 2015.

[20] M. Gonzalez-Franco, E. Ofek, Y. Pan, A. Antley, A. Steed, B. Spanlang, A. Maselli, D. Banakou, N. Pelechano, S. Orts-Escolano, et al. The rocketbox library and the utility of freely available rigged avatars. *Frontiers in virtual reality*, p. 20, 2020.

[21] J. W. Gurera and D. M. Isaacowitz. Emotion regulation and emotion perception in aging: A perspective on age-related differences and similarities. *Progress in Brain Research*, 247:329–351, 2019.

[22] A. Hanjalic. Extracting moods from pictures and sounds: Towards truly personalized tv. *IEEE Signal Processing Magazine*, 23(2):90–100, 2006.

[23] B. S. Hasler and D. A. Friedman. Sociocultural conventions in avatar-mediated nonverbal communication: A cross-cultural analysis of virtual proxemics. *Journal of Intercultural Communication Research*, 41(3):238–259, 2012.

[24] U. Hess, J. Dietrich, K. Kafetsios, S. Elkabetz, and S. Hareli. The bidirectional influence of emotion expressions and context: emotion expressions, situational information and real-world knowledge combine to inform observers' judgments of both the emotion expressions and the situation. *Cognition and Emotion*, 34(3):539–552, 2020.

[25] R. Hortensius, F. Hekele, and E. S. Cross. The perception of emotion in artificial agents. *IEEE Transactions on Cognitive and Developmental Systems*, 10(4):852–864, 2018.

[26] International Data Corporation. Ar/vr headset shipments grew dramatically in 2021, thanks largely to meta's strong quest 2 volumes, with growth forecast to continue, according to idc. `https://www.idc.com/getdoc.jsp?containerId=prUS48969722`, March 21 2022.

[27] C. C. Joyal, L. Jacob, M.-H. Cigna, J.-P. Guay, and P. Renaud. Virtual faces expressing emotions: an initial concomitant and construct validity study. *Frontiers in human neuroscience*, 8:787, 2014.

[28] D. Keltner and J. Haidt. Social functions of emotions at four levels of analysis. *Cognition & Emotion*, 13(5):505–521, 1999.

[29] P. Khooshabeh, P. J. Carnevale, C. de Melo, B. Volkman, J. Gratch, and J. Blascovich. Negotiation strategies with incongruent facial expressions of emotion cause cardiovascular threat. Technical report, 2013.

[30] M. Klasen, Y.-H. Chen, and K. Mathiak. Multisensory emotions: perception, combination and underlying neural processes. *Reviews in the Neurosciences*, 23(4):381–392, 2012.

[31] J. Lange, M. W. Heerdink, and G. A. Van Kleef. Reading emotions, reading people: Emotion perception and inferences drawn from perceived emotions. *Current opinion in psychology*, 43:85–90, 2022.

[32] 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*, pp. 1–10, 2017.

[33] P. Laukka and H. A. Elfenbein. Cross-cultural emotion recognition and in-group advantage in vocal expression: A meta-analysis. *Emotion Review*, 13(1):3–11, 2021.

[34] R. W. Levenson. The intrapersonal functions of emotion. *Cognition & Emotion*, 13(5):481–504, 1999.

[35] D. Maloney, G. Freeman, and D. Y. Wohn. " talking without a voice" understanding non-verbal communication in social virtual reality. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2):1–25, 2020.

[36] G. J. McHugo, J. T. Lanzetta, D. G. Sullivan, R. D. Masters, and B. G. Englis. Emotional reactions to a political leader's expressive displays. *Journal of personality and social psychology*, 49(6):1513, 1985.

[37] Meta Quest. Viseme Reference Images by Meta Quest. `https://developer.oculus.com/documentation/unity/audio-ovrlipsync-viseme-reference/`.

[38] J. M. Montepare and H. Dobish. The contribution of emotion perceptions and their overgeneralizations to trait impressions. *Journal of Nonverbal behavior*, 27(4):237–254, 2003.

[39] M. Mori, K. F. MacDorman, and N. Kageki. The uncanny valley [from the field]. *IEEE Robotics & automation magazine*, 19(2):98–100, 2012.

[40] E. Mower, S. Lee, M. J. Mataric, and S. Narayanan. Human perception of synthetic character emotions in the presence of conflicting and congruent vocal and facial expressions. In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 2201–2204. IEEE, 2008.

[41] E. Mower, M. J. Mataric, and S. Narayanan. Human perception of audio-visual synthetic character emotion expression in the presence of ambiguous and conflicting information. *IEEE Transactions on Multimedia*, 11(5):843–855, 2009.

[42] Oculus Horizon Worlds. Create. Explore. Together. `https://www.oculus.com/horizon-worlds/`.

[43] D. Polani. Emotionless chatbots are taking over customer service – and it's bad news for consumers. https://theconversation.com/emotionless-chatbots-are-taking-over-customer-service-and-its-bad-news-for-consumers-82962.

[44] S. D. Pugh. Service with a smile: Emotional contagion in the service encounter. *Academy of management journal*, 44(5):1018–1027, 2001.

[45] L. Rashotte. *Social Influence*. John Wiley Sons, Ltd, 2007.

[46] Rogo Digital LipSync. A powerful lip-syncing and facial animation tool for Unity. `https://lipsync.rogodigital.com/`.

[47] J. P. Scharlemann, C. C. Eckel, A. Kacelnik, and R. K. Wilson. The value of a smile: Game theory with a human face. *Journal of Economic Psychology*, 22(5):617–640, 2001.

[48] K. R. Scherer. Toward a dynamic theory of emotion: The component process model of affective states. *Geneva studies in Emotion and Communication*, 1:1–98, 1987.

[49] A. Schirmer and R. Adolphs. Emotion perception from face, voice,

and touch: comparisons and convergence. *Trends in cognitive sciences*, 21(3):216–228, 2017.

[50] C. Stanton and C. J. Stevens. Robot pressure: the impact of robot eye gaze and lifelike bodily movements upon decision-making and trust. In *International conference on social robotics*, pp. 330–339. Springer, 2014.

[51] T. J. Tanenbaum, N. Hartoonian, and J. Bryan. " how do i make this thing smile?" an inventory of expressive nonverbal communication in commercial social virtual reality platforms. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2020.

[52] K. L. Tidd and J. S. Lockard. Monetary significance of the affiliative smile: A case for reciprocal altruism. *Bulletin of the Psychonomic Society*, 11(6):344–346, 1978.

[53] I. Torre, E. Carrigan, K. McCabe, R. McDonnell, and N. Harte. Survival at the museum: A cooperation experiment with emotionally expressive virtual characters. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction*, pp. 423–427, 2018.

[54] I. Torre, E. Carrigan, R. McDonnell, K. Domijan, K. McCabe, and N. Harte. The effect of multimodal emotional expression and agent appearance on trust in human-agent interaction. In *Motion, Interaction and Games*, pp. 1–6. 2019.

[55] I. Torre, J. Goslin, and L. White. If your device could smile: People trust happy-sounding artificial agents more. *Computers in Human Behavior*, 105:106215, 2020.

[56] I. Torre, J. Goslin, L. White, and D. Zanatto. Trust in artificial voices: A" congruency effect" of first impressions and behavioural experience. In *Proceedings of the Technology, Mind, and Society*, pp. 1–6. 2018.

[57] G. A. Van Kleef, C. K. De Dreu, and A. S. Manstead. An interpersonal approach to emotion in social decision making: The emotions as social information model. In *Advances in experimental social psychology*, vol. 42, pp. 45–96. Elsevier, 2010.

[58] G. A. Van Kleef, A. C. Homan, B. Beersma, and D. van Knippenberg. On angry leaders and agreeable followers: How leaders' emotions and followers' personalities shape motivation and team performance. *Psychological Science*, 21(12):1827–1834, 2010.

[59] Z. Wang, H. Mao, Y. J. Li, and F. Liu. Smile big or not? effects of smile intensity on perceptions of warmth and competence. *Journal of Consumer Research*, 43(5):787–805, 2017.

[60] Z. Witkower and J. L. Tracy. Bodily communication of emotion: Evidence for extrafacial behavioral expressions and available coding systems. *Emotion Review*, 11(2):184–193, 2019.

[61] E. Zell, C. Aliaga, A. Jarabo, K. Zibrek, D. Gutierrez, R. McDonnell, and M. Botsch. To stylize or not to stylize? the effect of shape and material stylization on the perception of computer-generated faces. *ACM Transactions on Graphics (TOG)*, 34(6):1–12, 2015.

[62] K. Zibrek and R. McDonnell. Does render style affect perception of personality in virtual humans? In *Proceedings of the ACM symposium on applied perception*, pp. 111–115, 2014.