Raise Your Eyebrows Higher: Facilitating Emotional Communication in Social Virtual Reality Through Region-Specific Facial Expression Exaggeration

Xueyang Wang
Institute for Network Sciences and Cyberspace
Tsinghua University
Beijing, China
Zhongguancun Laboratory
Beijing, China
wang-xy22@mails.tsinghua.edu.cn

Sheng Zhao
Weiyang College
Tsinghua University
Beijing, China
zhao-s21@mails.tsinghua.edu.cn

Yihe Wang
Department of Computer Science and Technology
Tsinghua University
Beijing, China
wangyihe22@mails.tsinghua.edu.cn

Howard Ziyu Han
Human-Computer Interaction Institute
Carnegie Mellon University
pittsburgh, Pennsylvania, USA
ziyuh@andrew.cmu.edu

Xinge Liu
Department of Computer Science and Technology
Tsinghua University
Beijing, China
Beijing National Research Center for Information Science and Technology
Beijing, China
lxge@mail.tsinghua.edu.cn

Xin Yi*
Institute for Network Sciences and Cyberspace
Tsinghua University
Beijing, China
Zhongguancun Laboratory
Beijing, China
yixin@tsinghua.edu.cn

Xin Tong
Thrust of Computational Media and Arts
Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China
xint@hkust-gz.edu.cn

Hewu Li
Institute for Network Sciences and Cyberspace
Tsinghua University
Beijing, China
Zhongguancun Laboratory
Beijing, China
lihewu@cernet.edu.cn

Abstract

While exaggerated facial expressions in cartoon avatars can enhance emotional communication in social virtual reality (VR), they risk triggering the uncanny valley effect. Our research reveals that this effect varies significantly across different emotions. In Study 1 (N=30), participants evaluated scaled facial expressions during simulated VR conversations. We found that expression exaggeration had opposing effects: it decreased facial realism for joy, surprise, and disgust due to overly dramatic mouth movements, while enhancing realism for fear, sadness, and anger—emotions that rely on upper facial expressions typically constrained by HMD pressure. Based on these findings, we developed a region-specific facial expression exaggeration strategy that enhances under-expressed upper facial features while maintaining natural lower facial movements. Study 2 (N=20) validated this approach, demonstrating enhanced emotional intensity and contagion for negative emotions while mitigating the uncanny valley effect. Our research provides practical guidelines for optimizing avatar-mediated emotional communication in social VR environments.

*Corresponding author.

CHI '25, Yokohama, Japan

ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713688

CCS Concepts

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

Keywords

Social Virtual Reality, Facial Expression Exaggeration, Cartoon Avatars, User Empathy, Uncanny Valley Effect

ACM Reference Format:
Xueyang Wang, Sheng Zhao, Yihe Wang, Howard Ziyu Han, Xinge Liu, Xin Yi, Xin Tong, and Hewu Li. 2025. Raise Your Eyebrows Higher: Facilitating Emotional Communication in Social Virtual Reality Through Region-Specific Facial Expression Exaggeration. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 22 pages. https://doi.org/10.1145/3706598.3713688

1 Introduction

Social Virtual Reality (VR) has emerged as a leading application for immersive social interactions [69], enabling users to engage through head-mounted displays (HMDs) with avatars that emulate face-to-face communication [31, 79]. Unlike traditional avatar systems, social VR facilitates a comprehensive range of verbal and non-verbal interactions [75, 106].

Facial expressions play a crucial role in conveying emotional states in virtual environments [49, 120], fundamentally contributing to building meaningful social connections [73, 113]. Modern HMDs (such as Meta Quest Pro, Apple Vision Pro, and Pico 4 Pro) incorporate real-time facial tracking, translating detected movements into cartoon avatar expressions using the Facial Action Coding System [25]. However, the complexity of human expressions [20] and limitations in the fidelity of avatar models often result in compromised authenticity [33, 72], leading to lower ratings for intensity, arousal, and valence compared to human expressions [54, 81].

To address these challenges, researchers have explored facial expression exaggeration in cartoon avatars. Studies have shown that exaggerated expressions can enhance emotional communication and social presence [34, 87]. Interestingly, both exaggerated and dampened facial motions have been found to potentially improve avatar likability [47], suggesting a complex relationship between expression intensity and user perception.

While facial expression scaling shows potential for enhancing avatar emotional expression, it also carries risks, particularly the uncanny valley effect [48, 74]. This phenomenon can disrupt communication flow, reduce empathy, and diminish the sense of presence in social VR [72, 110]. Despite valuable insights from previous studies, a significant gap remains in understanding the impact of scaling real-time HMD-tracked expressions on emotional communication in social VR. Existing research has largely focused on static facial expressions [34, 47, 74, 93] or covers limited diversity of emotions [48, 87], failing to capture the dynamic, multi-emotional nature of facial expressions in social VR interactions.

To address these limitations, our study aims to answer two key research questions:

RQ1: *How does virtual avatar's facial expression scaling influence users' empathy for others' emotions and the uncanny valley effect in social VR?*

RQ2: *How can we leverage virtual avatar's facial expression scaling to enhance emotional communication in social VR while mitigating potential uncanny valley effects?*

We conducted two studies using pre-prepared materials to investigate facial expression scaling in social VR. Study 1 (N=30) examined the effects of expression scaling on user empathy and the uncanny valley effect across six basic emotions [26]. Surprisingly, our findings revealed that expression exaggeration did not significantly impact emotion recognition accuracy, highlighting the robustness of multimodal emotion processing in VR. More importantly, we uncovered a novel phenomenon: significant interaction effects between scaling and emotion categories on uncanny valley indicators. Exaggerated expressions for joy, surprise, and disgust increased eeriness and decreased appeal, while for fear, sadness, and anger, exaggeration enhanced realism and appeal.

In-depth analysis of user evaluations and facial movement patterns revealed that this phenomenon stems from inherent differences in facial expressions across emotions and physical limitations of HMD devices. As Figure 1 illustrates, emotions like joy, characterized by pronounced mouth and cheek movements, become visually unsettling when exaggerated further. Conversely, emotions such as anger, typically expressed through subtle upper facial movements which are often constrained by HMD pressure [123], benefit from exaggeration, enhancing their perceived naturalness and realism. This discovery challenges the notion of uniform scaling effects [47, 74], emphasizing the need for a nuanced approach to avatar expression enhancement.

Based on these insights, we proposed a region-specific facial expression exaggeration strategy, applying higher levels of exaggeration to upper-face regions while maintaining natural lower-face movements. Study 2 (N=20) validated the effectiveness of our strategy, demonstrating significant improvements in perceived emotional intensity and contagion while simultaneously mitigating the uncanny valley effect for fear, disgust, and anger. For joy, surprise, and sadness, the strategy maintained neutral or slightly positive impacts, avoiding potential negative effects of exaggeration. These results underscore the necessity of a nuanced, region-specific approach over a one-size-fits-all solution for enhancing emotional conveyance in social VR.

Our study makes three key contributions: (1) We uncover, for the first time, a significant interaction effect between avatar facial expression scaling and conveyed emotion on the uncanny valley effect in social VR, challenging the prevailing assumption of uniform impacts from expression scaling, (2) We propose and empirically validate a novel region-specific facial expression exaggeration strategy, offering a nuanced approach to enhance emotional expressiveness while mitigating the uncanny valley effect, and (3) We provide evidence-based design considerations for improving emotional communication in social VR.

2 Related Work

In this section, we review relevant literature and highlight how our work addresses existing research gaps, contextualizing our study within the broader field of avatar-based interactions in virtual environments.

2.1 Enhancing Avatar Facial Expressions in Social VR

In avatar-mediated communication, facial expressions are crucial for conveying and understanding emotional states, facilitating effective interaction [94]. Dynamic full-face expressions, particularly eye and mouth movements, significantly enhance social presence in collaborative virtual tasks [55]. Synchronous mapping of facial tracking onto virtual faces induces high levels of face-ownership and agency [58], while integrated eye and face tracking benefits certain enfacement illusion scales [36]. Even in mixed reality settings, virtual facial expressions improve collaboration quality and social presence [14]. Moreover, increased animation realism through real-time tracking enhances the perceived appeal of virtual faces [58], with the authenticity and fidelity of facial appearance significantly impacting social communication in virtual environments [22, 52].

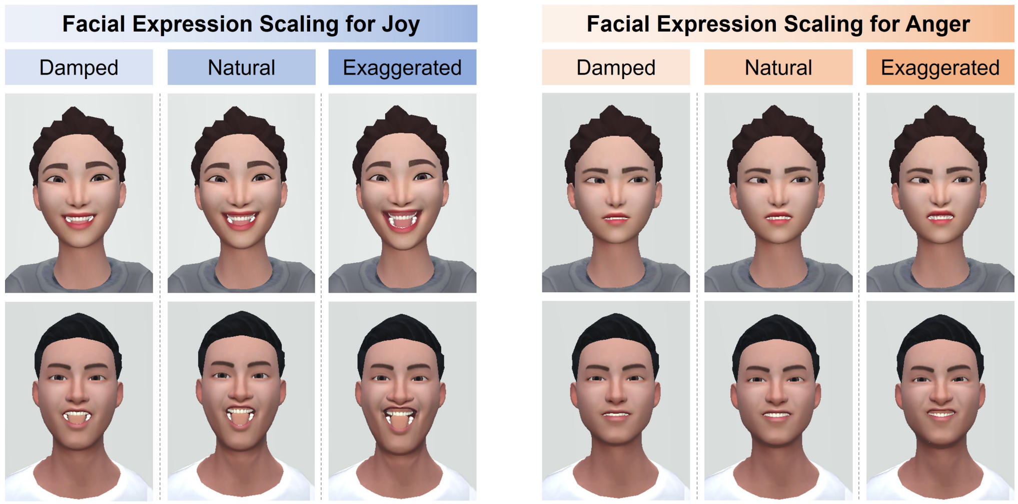


Figure 1: Illustrations of facial expression scaling on virtual avatars for Joy and Anger, demonstrating three levels: Damped (reduced amplitude), Natural (accurately captured amplitude), and Exaggerated (enhanced amplitude). Left: Joy expressions are naturally pronounced, particularly in the mouth region, and can appear excessively intense when exaggerated. Right: Anger expressions, where upper facial features (eyebrows, eyelids, and cheeks) are often subtle and barely perceptible in their natural state, become more vivid and recognizable when exaggerated. These representative moments highlight how exaggeration can either diminish or enhance the perceived realism of the expression depending on the emotion portrayed.

Despite the importance of facial expressions in social communication, tracking them in XR environments presents unique challenges due to HMD obstruction. Researchers have addressed this by employing custom hardware or integrating existing sensors, often combining lower face tracking with eye-tracking capabilities [68, 88, 100]. Recent advancements have led to commercial HMDs supporting real-time full facial tracking, such as Meta Quest Pro, Apple Vision Pro, Pico 4 Pro, and VIVE Full Face Tracker, which estimate facial movements using inward-facing cameras. The widely adopted Natural Facial Expressions approach converts detected facial movements into activations based on the Facial Action Coding System (FACS), mapping these to artist-created blendshapes representing avatar facial expressions.

However, accurately mirroring the intricate details of human facial expressions in avatars remains challenging [20]. The limited model granularity of cartoon avatars and simplifications in FACS-based muscle movement mapping [82] often compromise authenticity. Subtle expressions may be overlooked, resulting in the loss of prosocial cues [33, 61, 72]. Additionally, HMD pressure on facial skin and muscles can alter users' expression patterns [11]. Facial expressions during muscle strain are reported to be characterized by tension of the corrugator supercilii and zygomaticus major muscles [123]. These limitations in behavioral realism can affect users' perception of emotions, potentially impacting communication quality [38]. Consequently, virtual facial expressions often receive lower ratings for intensity, arousal, and valence compared to human expressions [54, 81].

To improve avatar expression realism and enhance social interaction quality, researchers have explored various enhancement techniques. Facial expression scaling, which modulates the intensity of facial movements, shows promise. For cartoon-style avatars, exaggerated facial movements are widely used to enhance emotional cues [47]. Studies suggest that augmenting facial expressions could improve social VR experiences [34], with enhanced smiles leading to more positive affect and stronger social presence [87]. However, a significant research gap exists regarding the effects of systematic expression scaling on user empathy, perceived authenticity, and the uncanny valley effect in social VR, particularly for real-time tracked faces in immersive environments.

Recent cutting-edge work has also explored using affective computing and computer graphics methods to optimize avatar dynamic expressions. Kang et al. identified emotion-based prioritized facial expressions that are crucial for avatar-mediated communication [51]. Some approaches use multimodal information such as speech [91, 111] or reference facial videos [4, 119] to identify emotions and generate more appropriate expressions. Machine learning methods have also been employed to enhance geometric consistency and perceptual validity of avatar expressions [90, 92]. However, due to the developmental nature of these technologies and their high computational requirements, they are currently more prevalent in computational animation and have not yet been widely adopted in Social VR applications.

2.2 User Empathy for Virtual Avatar

Empathy in virtual environments mirrors real-world interactions, relying on multiple channels. Schirmer and Adolphs [108] emphasize how facial, vocal, and tactile signals converge to form a holistic understanding of emotions. In virtual contexts, facial expressions serve as a primary channel for conveying both the emotional state and behavioral intentions of an individual, playing a fundamental role in non-verbal communication [24, 27]. Beyond facial muscle movements, subtle eye-related features such as fixations and pupil size variations also convey rich emotional information [59, 65].

Research demonstrates that humans can accurately recognize facial expressions of both humanoid and non-humanoid virtual agents, comparable to real human expressions [23, 37, 43]. However, representing human emotions using 3D virtual characters has inherent limitations due to reduced precision and fidelity [111]. Factors influencing emotion perception in virtual environments include interpersonal distance [17], visual realism [40], expression driving methods [44, 45, 118], avatar anthropomorphism [114], personalization [96], and character type [5]. Additional visual cues, such as biosignal visualizations, have been shown to enhance emotion perception and empathy towards avatars [66, 122].

Voice also plays a crucial role in empathy within virtual environments [64, 112], conveying information beyond semantics [3]. Some studies have explored user perceptions when emotions conveyed through different channels are inconsistent [13, 124]. These researches are particularly important for social VR, where users naturally integrate facial expressions, speech, voice, lip movements, head motions, and overall appearance for empathy in interpersonal interactions. However, there is a lack of research analyzing whether real-time tracked facial expressions in virtual environments align with emotions conveyed through voice and posture in terms of type and intensity.

Psychological research indicates that some emotions are better recognized than others interpersonally [29], and similar expressions are often confused [102, 103]. This variation in emotion recognition accuracy has also been observed with virtual avatars. For instance, Geraets et al. found lower accuracy rates for disgust and happiness in static facial expressions in VR [32]. Park et al.'s study on Memojis revealed that happiness and sadness are most robustly conveyed, while fear and disgust had low levels of conveyance, and anger, contempt, and surprise were relatively inaccurately perceived [93, 93]. However, there is a lack of research comparing the expressiveness of real-time facial-tracked avatar expressions across different emotion categories in VR.

2.3 Mitigating the Uncanny Valley Effect in Social VR

The uncanny valley effect, introduced by Mori in 1970 [83], describes a phenomenon where affinity towards humanoid figures initially increases with human likeness, then sharply decreases at a certain point, before rising again as the figure becomes nearly indistinguishable from a real human [78]. This effect significantly impacts social interactions in virtual environments, potentially disrupting user engagement and emotional connection [110].

In social VR contexts, avatar realism has been identified as a key factor contributing to the uncanny valley effect. Highly realistic avatars may elicit stronger negative reactions [62, 72], while cartoon-like characters often prove more attractive and manageable in terms of facial expressions. Researchers acknowledge that the uncanny valley is related to both avatar appearance and behavior [9, 53], suggesting that combining appearance and motion fidelity is crucial for achieving a balanced avatar design. Notably, the motion fidelity of the face's upper part appears to be a critical aspect [117, 126].

Facial expression scaling, while potentially intensifying emotional conveyance, presents a complex challenge that may trigger the uncanny valley effect. Hyde et al. [48] found that exaggerated facial expressions positively correlated with perceived extroversion and persuasiveness but negatively impacted perceived realism and naturalness. Mäkäräinen et al. [74] demonstrated that high exaggeration is suitable only for less realistic avatars, as increased realism can easily lead to the uncanny valley effect. Additionally, the intensity of emotions expressed by virtual humans influences uncanny valley ratings [40].

However, these studies primarily focus on animated characters presented on 2D screens. There is a lack of comprehensive research on whether scaling avatar expressions in immersive social VR environments triggers the uncanny valley effect. Furthermore, the relationship between the type of emotion expressed by users and the uncanny valley effect caused by avatars remains understudied. This gap in the literature highlights the need for further investigation into the nuanced interactions between emotion category, expression scaling, and the uncanny valley phenomenon in social VR contexts.

3 Collecting and Modeling the Avatar's Facial Expressions

To collect natural emotional expressions in social VR for our experimental stimuli, we developed a comprehensive material collection and processing methodology. Inspired by Kullmann et al. [61], our approach involved recording, scaling, and replaying facial expressions. As illustrated in the flowchart (Figure 2), our material preparation comprised several interconnected stages. We began by creating avatars and developing a Unity scene to simulate a realistic social VR environment. Trained actors then performed emotional expressions in this simulated VR setting based on provided scripts. Using HMDs, we captured their voice, facial expressions, eye movements, and body postures. Subsequently, we trimmed irrelevant sections such as blank periods and greetings in recorded clips. We also checked the data quality and accuracy of each clip. Finally, we scaled the facial expressions in these clips to create our experimental stimuli.

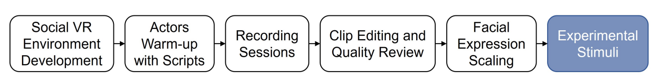


Figure 2: Flowchart illustrating the collection and proprocessing process for experimental stimulus in our study.

This collection process enables controlled manipulation of facial expressions while maintaining consistent content across conditions, crucial for isolating the effects of expression scaling and minimizing confounding variables. The following sections provide detailed descriptions of each stage in this process.

3.1 Collection Apparatus and Environment

To simulate a realistic and representative social VR scenario, we constructed a virtual environment mimicking an indoor friend-to-friend conversation. We chose a face-to-face dialogue setup, the

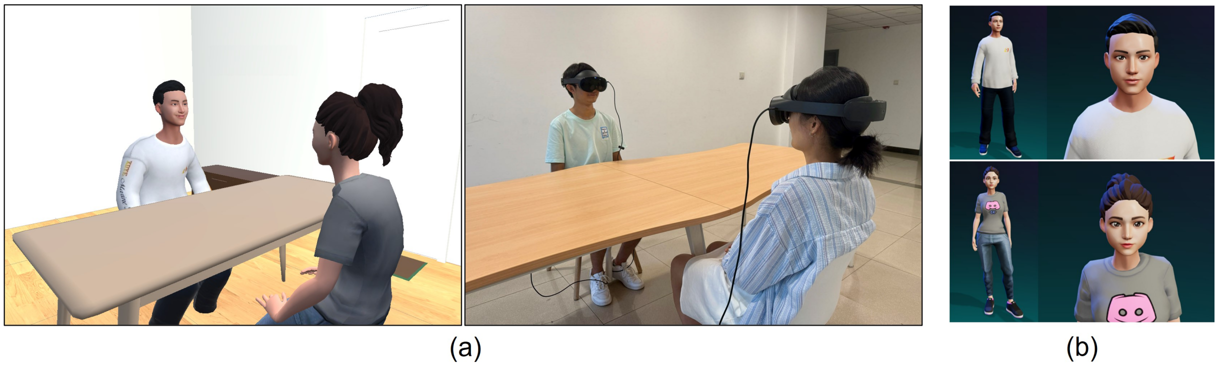


Figure 3: (a) The virtual (left) and real (right) environments for data collection; (b) The cartoon avatar used in our study.

most common social interaction mode in social VR [69]. As illustrated in Figure 3(a), during the experimental stimulus collection, the experimenter and actor sat across a table in both virtual (embodied in gender-matched avatars) and real environments. This setup allowed for real-time observation of each other's avatar facial expressions and body movements, along with voice communication, similar to platforms like VRChat.

We developed a custom Unity 2021.3.26f1 application running on ASUS laptops, streamed to Meta Quest Pro headsets. The Meta Quest Pro offers a 72Hz refresh rate, 1800 × 1920 resolution per eye, and a 106° horizontal by 96° vertical field of view. Our laptops ran Windows 11 with Intel i9-12900H CPUs, 32.0 GB of memory, and NVIDIA GeForce RTX 3080 Ti GPUs. This virtual environment and hardware setup remained consistent across stimulus collection and both studies.

We employed cartoon-style avatars in our experiments, aligning with the design choices of mainstream social VR platforms such as Meta Horizon Worlds[1], VRChat[2], Roblox[3], MeetinVR[4], and Rec Room[5]. This design enhances the generalizability of our findings to current social VR applications. The prevalence of cartoon avatars in social VR can be attributed to their ability to provide sufficient embodiment and visual appeal [72], while reducing uncanny valley effects [63]. Additionally, they offer greater customizability [1], ensure privacy protection [70], and are computationally efficient, making them well-suited for various hardware configurations.

Specifically, we created full-body avatar models using Ready Player Me[6], a cross-game platform used by major social VR applications. Full-body avatars were selected based on research demonstrating their superior performance over partial body configurations [120]. As shown in Figure 3(b), we created one male and one female generic avatar. A professional 3D animation artist added 70 facial blendshapes to each avatar, following Face Tracking API guidelines[7]. We implemented Unity scripts to map facial expressions detected by the HMD's inward-facing cameras to these blendshapes, enabling real-time facial tracking. Our application also utilized the Oculus Movement SDK for Unity for comprehensive body, face, and eye tracking.

3.2 Recruitment and Warm-up of Actors

To ensure authentic performances while minimizing individual biases, we recruited five pairs of trained actors (one male, one female per pair) through our university drama team instructor. The cohort comprised 10 Chinese actors, averaging 22 years of age (SD = 1.9), each with 2-3 years of training and stage experience. We screened and paired applicants, matching their demographics with avatar appearances to enhance realism. All actors provided written informed consent, and the study received approval from the university's ethics committee. The 75-minute collection process concluded with each actor receiving $20 compensation for their participation.

Following the method adopted by Kullmann et al. [61], actors performed according to simple scripts. In our scenario, actors narrated personal experiences to a friend (the experimenter) while conveying one of Ekman's six basic emotions [26]: Joy, Sadness, Fear, Anger, Surprise, and Disgust. These emotions were chosen for their universal recognition across cultures.

We crafted three distinct scripts for each of the six emotions, totaling 18 scripts. These varied in topic within each emotional category to reduce potential single-topic influence. All scripts were written in Chinese, matching the native language of our actors and study participants. Each script comprised a 30-second emotional monologue, designed to capture a full emotional arc while maintaining engagement. Content was tailored to reflect both the intended emotions and relatable university student experiences, ensuring authentic performances.

Following Yang et al.'s [124] approach, we generated initial scripts based on common online conversation topics. The research team then collaboratively reviewed and refined these to ensure appropriate length, accurate emotion conveyance, and content authenticity. For transparency and reproducibility, detailed English translations of all scripts are provided in the Supplemental Material.

The 18 scripts were randomly distributed among the actor pairs, with each actor performing nine scripts. Prior to the recording session, actors familiarized themselves with the VR environment

[1]https://horizon.meta.com
[2]https://hello.vrchat.com
[3]https://www.roblox.com
[4]https://www.meetinvr.com
[5]https://recroom.com
[6]https://readyplayer.me
[7]https://developer.oculus.com/documentation/unity/move-face-tracking

Table 1: The average activation levels of different facial blendshapes across various emotional expressions, with darker colors indicating higher strength. Symmetric facial blendshapes were averaged.

Facial region	Blendshapes	Joy	Surprise	Disgust	Fear	Sadness	Anger
Eyebrows	Brow Lowerer	5.18%	3.28%	7.09%	5.30%	3.93%	5.59%
	Inner Brow Raiser	2.93%	2.11%	2.42%	1.82%	0.91%	1.38%
	Outer Brow Raiser	1.57%	1.76%	1.60%	1.34%	0.45%	1.02%
Gaze	Eyes Closed	8.97%	7.90%	11.10%	8.14%	9.62%	9.36%
	Eyes Look Right	7.74%	8.56%	7.33%	6.60%	9.40%	7.96%
	Eyes Look Left	2.17%	3.33%	3.83%	7.79%	3.11%	3.27%
	Eyes Look Down	4.35%	2.27%	4.09%	2.52%	5.45%	3.27%
	Eyes Look Up	0.88%	2.86%	0.95%	1.96%	0.11%	0.72%
Eyelid	Lid Tightener	21.97%	11.86%	16.73%	8.63%	3.42%	10.49%
	Upper Lid Raiser	0.18%	0.46%	0.93%	0.99%	0.20%	0.69%
Cheeks	Cheek Raiser	26.79%	12.11%	11.89%	5.95%	4.89%	8.05%
	Cheek Puff	0.04%	0.04%	0.07%	0.06%	0.05%	0.04%
	Cheek Suck	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Nose	Nose Wrinkler	1.28%	0.82%	5.05%	1.96%	0.56%	2.12%
Mouth	Upper Lip Raiser	38.05%	25.47%	24.22%	13.01%	9.16%	16.66%
	Lip Corner Puller	36.45%	16.80%	9.49%	3.19%	1.69%	4.77%
	Dimpler	20.54%	12.16%	7.82%	4.04%	3.14%	5.73%
	Chin Raiser	9.17%	6.38%	6.89%	3.62%	2.82%	4.77%
	Lip Suck	9.17%	5.20%	3.88%	2.06%	1.63%	2.50%
	Lower Lip Depressor	11.63%	5.64%	3.01%	0.87%	0.80%	1.20%
	Lip Stretcher	7.84%	3.42%	2.32%	0.86%	1.06%	1.42%
	Lips Toward	2.41%	2.13%	2.38%	2.15%	1.95%	2.21%
	Lip Pucker	0.67%	1.21%	2.38%	1.94%	1.10%	1.71%
	Lip Tightener	0.15%	0.42%	0.64%	0.68%	0.49%	0.60%
	Mouth Left	0.73%	0.48%	0.51%	0.31%	0.17%	0.42%
	Lip Corner Depressor	0.43%	0.26%	0.58%	0.35%	0.39%	0.42%
	Lip Pressor	0.23%	0.15%	0.78%	0.27%	0.12%	0.40%
	Lip Funneler	0.08%	0.16%	0.19%	0.33%	0.22%	0.20%
	Mouth Right	0.05%	0.02%	0.03%	0.01%	0.02%	0.02%
Jaw	Jaw Drop	16.46%	13.57%	12.19%	10.24%	7.48%	9.76%
	Jaw Thrust	1.68%	2.28%	3.19%	2.20%	0.68%	2.45%
	Jaw Sideways Right	0.72%	0.44%	0.46%	0.26%	0.19%	0.31%
	Jaw Sideways Left	0.12%	0.13%	0.21%	0.14%	0.08%	0.12%

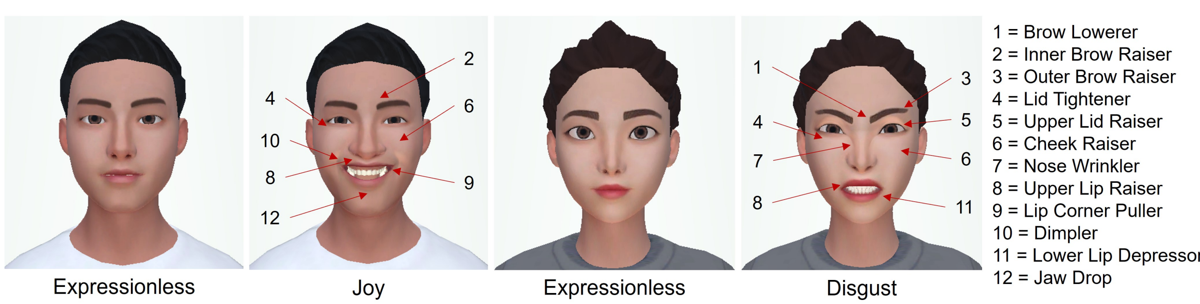


Figure 4: Examples of highly activated blendshapes in facial expressions for Joy and Disgust. Expressionless faces for comparison.

by wearing the HMD. They were given 3-5 minutes to warm up with each script and encouraged to rehearse briefly to ensure natural and accurate performances.

3.3 Collection Procedure and Post-processing

Prior to data collection, we calibrated the HMDs to prevent posture abnormalities, ensuring the accuracy of collected motion data. Recording sessions commenced with friendly greetings, followed by the actor's emotional narration of the scripted experience. To maintain conversational realism, the experimenter provided non-verbal feedback such as nodding and smiling. Throughout the session, we recorded the actors' voices, facial expressions, gaze directions, and body postures. After each performance, actors rated their emotional intensity on a 7-point Likert scale. A 3-minute interval between sessions allowed actors to reset emotionally, ensuring distinct emotional states for each recording.

We collected a total of 90 emotional expression clips from the five actor pairs, with an average duration of 32.2 seconds (SD = 8.3). These clips underwent editing to remove irrelevant sections such as blank periods and greetings, focusing on the core emotional content. We conducted thorough reviews to ensure data quality, accuracy, and integrity of each emotional expression. This comprehensive approach to data collection and preparation allowed us to create a robust dataset of natural, diverse emotional expressions, crucial for the validity of our subsequent analyses.

3.4 Overview of Collected Facial Expressions

The facial expressions in our study are categorized based on the Facial Action Coding System (FACS) [25]. Each expression is represented as a blendshape with an activation level ranging from 0% (completely relaxed) to 100% (maximum amplitude), indicating the intensity of the facial expression. Table 1 presents the average activation levels of facial blendshapes across various emotions, with symmetric blendshapes averaged. The high-fidelity facial tracking technology in the Meta Quest Pro headsets ensures precise capture of facial movements [18].

Our data reveal significant variations in blendshape activation across facial regions and emotions. Joy, Surprise, and Disgust expressions exhibited greater amplitude, particularly in the Eyelid, Cheeks, and Mouth regions. For instance, Joy showed high activation in *Cheek Raiser* (26.79%) and *Upper Lip Raiser* (38.05%), while Disgust had pronounced *Lid Tightener* (16.73%) and *Nose Wrinkler* (5.05%) activations. Surprise demonstrated notable *Jaw Drop* (13.57%) activation. These findings align with FACS-based studies [20, 57], which report larger facial movements for emotions involving smiling (Joy), mouth opening (Surprise), and wrinkling (Disgust). In contrast, Fear, Sadness, and Anger expressions were more subtle, characterized by micro-expressions [116]. For example, Sadness showed relatively low activation across most blendshapes, with its highest being *Eyes Closed* at 9.62%, while Anger's most prominent activation was in the *Upper Lip Raiser* at 16.66%.

To facilitate reader understanding of blendshape locations and movement characteristics, Figure 4 illustrates typical expressions for Joy and Disgust emotions, highlighting the blendshapes with the highest activation levels. This visual representation provides a clear depiction of how different emotions manifest through specific facial features and movements.

4 Study 1: Modeling the Effect of Facial Expression Scaling

We conducted a controlled VR experiment using our collected experimental stimuli. This study aimed to investigate how facial expression scaling influences users' empathy for others' emotions and the uncanny valley effect in social VR environments.

4.1 Study Design

We employed a 3×6 within-subjects design, where participants experienced a series of audiovisual stimuli in the same virtual environment used for data collection. These stimuli varied across three levels of *Expression Scaling* and six *Emotion Category*. An a priori power analysis [121] indicated that a sample of 30 participants would achieve a statistical power of 0.8, assuming a medium effect size (d = 0.5) and an alpha error probability of 0.05.

As illustrated in Figure 1, we defined three scaling levels:

- **Natural**: Original detected facial blendshape activation levels, serving as the baseline for comparison.
- **Damped**: 66.7% (2/3) of the original strength. This value was chosen to provide a noticeable reduction in expression intensity without completely flattening the emotion. It allows for subtle expressions while maintaining recognizability.
- **Exaggerated**: 150% of the original strength. This 1.5× amplification was selected to enhance expressiveness significantly without creating extreme distortions. It aims to intensify emotions while staying within a plausible range of human expression.

These values were determined through a combination of literature review [47, 87] and pilot testing to ensure discernible differences between conditions while avoiding excessive exaggeration that might cause discomfort or unnatural appearances. We adjusted all blendshapes except those related to gaze direction to avoid unnatural eye movements.

4.2 Measurements

To comprehensively assess the impact of scaled emotional expressions, participants completed a questionnaire after each recorded clip. Following the questionnaire, we conducted short interviews with open-ended questions to gather more nuanced, qualitative insights. The questionnaire and open-ended questions are summarized in Table 2 and fully available in the Supplementary Material.

Our evaluation of user empathy was based on Powell and Roberts' [95] conceptualization of empathic responses, focusing on two dimensions: (1) Cognitive Empathy: Assessed through emotion recognition accuracy and perceived differences in emotion intensity [56], measuring participants' ability to correctly identify and gauge the strength of expressed emotions; (2) Affective Empathy: Measured via the degree of emotional contagion experienced by participants [21], indicating how effectively the expresser's emotions were transmitted to the perceiver.

To evaluate the uncanny valley effect, we adopted scales widely used in previous studies [41, 42, 118]. These included measures of vividness, appeal, realism, and eeriness of facial expressions, as well as overall comfort. Recognizing the importance of multimodal consistency in virtual interactions [74], we also included a measure of perceived consistency between facial expressions and speech.

4.3 Participants and Procedure

We recruited 30 Chinese participants (15 females, 15 males) from a local university, aged 18 to 28 (M = 23.2, SD = 2.3). All participants were naive to the experiment's objectives and had normal or corrected-to-normal vision. Their familiarity with VR was assessed

Table 2: Dependent variables collected in the questionnaire and the short interview.

User Empathy	**Emotion Recognition Accuracy**: Recognition rate of six emotion categories (Joy, Sadness, Fear, Anger, Surprise, Disgust). **Emotional Intensity Difference**: Discrepancy between perceiver-assessed and expresser self-reported emotional intensity. **Emotional Contagion**: Perceiver's self-reported empathy level (1: untouched, 7: highly empathetic).
Uncanny Valley Effect	**Expression-Speech Consistency**: Perceived alignment of facial expressions with speech (1: inconsistent, 7: consistent). **Facial Vividness**: Assessing the richness and clarity of facial expressions (1: bland and rigid, 7: rich and vivid). **Facial Appeal**: Measuring the attractiveness or likeability of facial expressions (1: unappealing, 7: appealing). **Facial Realism**: Gauging how true-to-life the expressions appeared (1: artificial, 7: realistic). **Facial Eeriness**: Evaluating any unsettling or disconcerting feelings evoked (1: normal, 7: eerie). **Comfort**: Assessing overall user comfort with the interaction (1: uncomfortable, 7: comfortable).
Open-ended Questions	**Emotion Recognition Cues**: Key indicators used for emotion recognition. **Expression-Speech Discrepancies**: Noted misalignments between facial expressions and speech. **Facial Expression Realism**: Factors contributing to perceived realism or lack thereof. **Facial Expression Appeal**: Elements affecting attractiveness of expressions. **Comfort Factors**: Reasons behind comfort or discomfort during the experience.

Table 3: Description of the coding scheme identified in the thematic analysis of the qualitative data from the open-ended answers.

Theme	Category	Description
Emotion Recognition	Modalities	Communication channels (e.g., facial expressions, gaze, body movements, speech).
	Perceived Indicators	Observed emotional cues (e.g., melancholic voice, smile, bowed head).
Facial Features	Facial Region	Specific facial areas mentioned (e.g., eyebrows, cheeks, lips).
	Movements	Facial movements described (e.g., raise, open, frown).
	Scale	Extent of facial movements (e.g., insufficient, appropriate, excessive).
	Inconsistency	Misalignments between facial features and speech (e.g., lip sync, rhythm, emotion).
Assessments	Positive	Favorable descriptions (e.g., natural, appropriate, vivid).
	Negative	Unfavorable descriptions (e.g., uncanny, rigid, bland).

on a 5-point scale (M = 2.97, SD = 1.13), with six participants reporting prior experience with social VR applications like VRChat. Participants provided written informed consent, and the study was approved by the university's ethics committee. The entire experiment lasted approximately 90 minutes, with participants receiving $20 compensation for their time and effort.

Prior to the experiment, participants were given 5 minutes to familiarize themselves with the equipment and virtual environment. They were instructed to regard the expresser as a close friend sharing personal experiences in a virtual face-to-face conversation. Participants were asked to focus on perceiving their "friend's" emotions through observation and listening.

The experiment consisted of 18 clips, with questionnaires and short interviews completed after each. To control for gender effects, clips were drawn from the same actor pair and presented using gender-matched avatars. The 30 participants were divided into five groups, each experiencing expressions from different actor pairs. We ensured equal probability of assignment for each emotion across the three scaling conditions. The order of scaling levels and emotion categories was counterbalanced to mitigate order effects. A 1-minute break was provided between clips to prevent fatigue and refresh emotional states.

4.4 Analysis

We analyzed 540 questionnaire responses (18×30) with no outliers removed. Due to non-normal distribution, we employed a two-way repeated measures ANOVA with the Aligned Rank Transform (ART) [30] for non-parametric factorial analysis ($\alpha = 0.05$). Scaling Level and Emotion Category served as within-subject factors, with three user empathy measures and six uncanny valley effect indicators as dependent variables. We applied Mauchly's test for sphericity [76] and Greenhouse-Geisser corrections where necessary. Post-hoc comparisons utilized Bonferroni-corrected paired t-tests, reporting effect sizes as partial eta-squared (η_p^2).

Table 4: Summary of quantitative results. NS denotes nonsignificant results, while asterisk (*) indicates significant results (* at the < .05 level, ** at the < .01 level, and * at the < .001 level). Values in parentheses represent pairwise comparison results: D stands for *Damped*, N stands for *Natural*, and E stands for *Exaggerated*. Jo, Sa, Fe, An, Su, and Di refer to *Joy*, *Sadness*, *Fear*, *Anger*, *Surprise*, and *Disgust*, respectively.**

Measurements		Main Effect of Expression Scaling	Main Effect of Emotion Category	Interaction Effect
User Empathy	Emotion Recognition Accuracy	NS	*** (Su, Di, Fe, An < Jo; Di, Fe, An < Sa)	NS
	Emotional Intensity Difference	NS	**	NS
	Emotional Contagion	NS	** (Fe, Su < Di)	NS
Uncanny Valley Effect	Expression-Speech Consistency	NS	*** (Su, Fe, Sa, An < Jo; Su, Fe < Di)	NS
	Facial Vividness	** (D < E)	*** (Fe, Sa, An < Jo; Fe, Sa, An < Di)	NS
	Facial Appeal	NS	** (Su < Di)	**
	Facial Realism	* (E < N)	* (Su < Di)	**
	Facial Eeriness	*** (E < D, N)	*** (Di, Sa, An < Su, Sa < Jo)	***
	Comfort	** (E < D, N)	NS	**

The open-ended responses from brief interviews were recorded and transcribed, providing rich, contextual data to complement the quantitative measures. We conducted a hybrid thematic analysis, combining elements of both inductive and deductive approaches [115]. Initially, we used inductive thematic content analysis [77] with incremental open coding [15] to identify emergent themes. The research team collaboratively developed an initial coding scheme (Table 3) based on these themes. Then, two researchers deductively coded the data using the scheme, resolving disagreements through discussion. We calculated inter-coder reliability, achieving a Cohen's Kappa of 0.74, indicating good consistency between the two coders.

4.5 Quantitative Results

Table 4 summarizes the main statistical findings of our study, which reveal complex relationships between facial expression scaling, emotion category, and user perceptions in social VR contexts. Key results include:

- *Expression Scaling* did not significantly influence user empathy.
- *Joy* and *Sadness* exhibited higher recognition accuracy, while *Disgust* more effectively triggered emotional contagion.
- A significant interaction effect was observed between *Expression Scaling* and *Emotional Category* on uncanny valley indicators (facial appeal, realism, eeriness, and comfort).
- The main effect of *Expression Scaling* on the uncanny valley effect: *Exaggerated* expressions increased vividness but decreased facial realism, increased eeriness, and lowered comfort.
- The main effect of *Emotion Category* on the uncanny valley effect: *Joy* and *Disgust* showed higher expression-speech consistency and facial vividness; *Disgust* exhibited higher facial realism and appeal compared to surprise; *Surprise* and *Joy* induced higher eeriness.

These findings highlight the nuanced impact of facial expression scaling on user experience in social VR, suggesting that the effects are not uniform across different emotions. The following sections provide detailed statistical analyses for each aspect of our study.

4.5.1 User Empathy. *Expression Scaling* has no significant effect on user empathy (Figure 5). However, despite the lack of statistical significance, mean values suggest that increasing the expression scaling slightly enhances emotion recognition accuracy, from *Damped* ($M = 0.83, SD = 0.15$) to *Natural* ($M = 0.84, SD = 0.11$) to *Exaggerated* ($M = 0.86, SD = 0.15$). Similarly, for Emotional Intensity Difference, higher expression scaling leads perceivers to judge the expresser's emotional intensity more strongly, from *Damped* ($M = 0.16, SD = 0.17$) to *Natural* ($M = 0.33, SD = 0.19$) to *Exaggerated* ($M = 0.51, SD = 0.14$). As for Emotional Contagion, there is virtually no difference among *Damped* ($M = 4.99, SD = 0.16$), *Natural* ($M = 4.97, SD = 0.16$), and *Exaggerated* ($M = 4.96, SD = 0.17$). These results suggest that while expression scaling may subtly influence emotion recognition and perceived intensity, its impact on overall empathy in social VR is limited.

In contrast, *Emotion Category* has a significant impact on user empathy (Figure 6). Regarding Emotion Recognition Accuracy, the recognition rate for *Joy* ($M = 1.0, SD = 0.0$) is significantly higher than for *Surprise* ($M = 0.83, SD = 0.27, p < .05$), *Disgust* ($M = 0.81, SD = 0.24, p < .01$), *Fear* ($M = 0.73, SD = 0.22, p < .0001$), and *Anger* ($M = 0.72, SD = 0.26, p < .0001$), whereas the recognition rate for *Sadness* ($M = 0.98, SD = 0.08$) is significantly higher than for *Disgust* ($p < .05$), *Fear* ($p < .0001$), and *Anger* ($p < .0001$). This pattern is consistent with previous studies [8, 40, 86, 93, 127], which have found that happiness and sadness are more easily recognized in both real and cartoon faces. The higher recognition accuracy for these emotions attributes to their distinct and culturally universal expression patterns.

For Emotional Intensity Difference, no significant pairwise differences were observed across emotions. However, in terms of

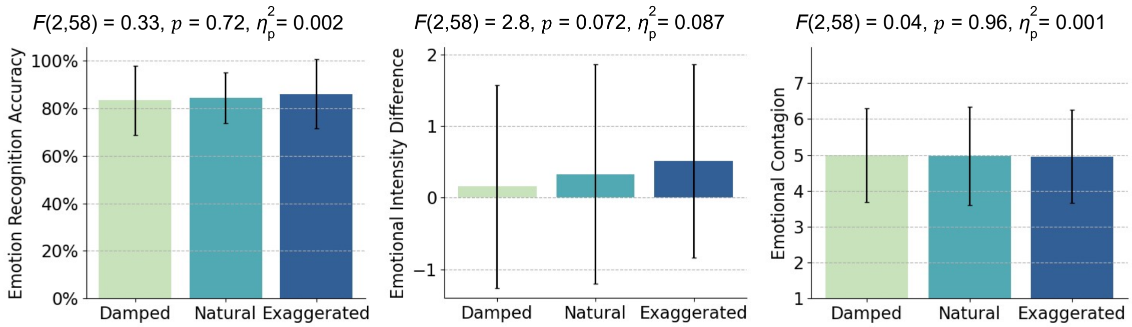


Figure 5: The main effect of *Expression Scaling* on Emotion Recognition Accuracy (left), Emotional Intensity Difference (middle), and Emotional Contagion (right). The error bars denote standard deviation.

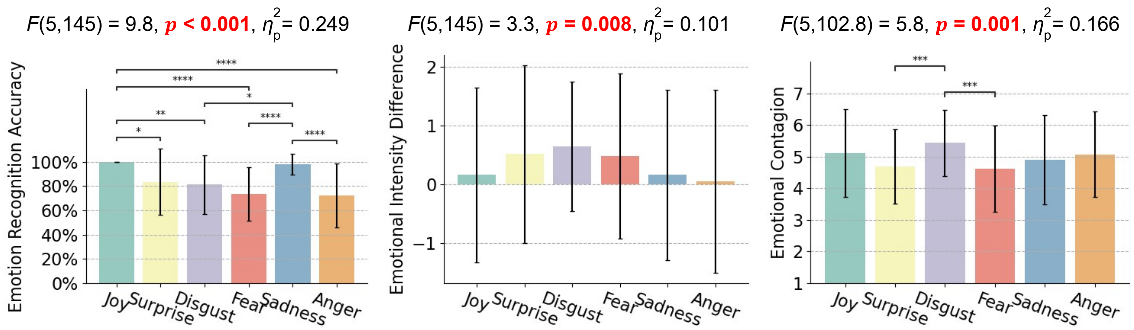
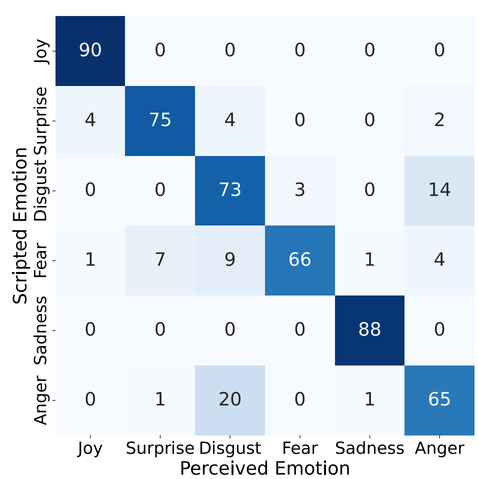


Figure 6: The main effect of *Emotion Category* on Emotion Recognition Accuracy (left), Emotional Intensity Difference (middle), and Emotional Contagion (right). The error bars denote standard deviation, and *, **, *, and **** denote significant differences at $p < 0.05$, $p < 0.01$, $p < 0.001$, and $p < 0.0001$ levels, respectively.**



	Joy	Surprise	Disgust	Fear	Sadness	Anger
Joy	90	0	0	0	0	0
Surprise	4	75	4	0	0	2
Disgust	0	0	73	3	0	14
Fear	1	7	9	66	1	4
Sadness	0	0	0	0	88	0
Anger	0	1	20	0	1	65

Figure 7: Confusion matrix illustrating emotion recognition accuracy for six basic emotions, with the three expression scaling levels considered together.

Emotional Contagion, *Disgust* ($M = 5.43, SD = 0.16$) scored significantly higher than both *Fear* ($M = 4.61, SD = 0.20, p < .001$) and *Surprise* ($M = 4.70, SD = 0.16, p < .001$). We hypothesize that this heightened emotional contagion for *Disgust* may be attributed to its more pronounced facial expressions, particularly in the eyebrows, eyelid, and nose regions (Table 1). This aligns with research suggesting that disgust expressions can elicit strong emotional contagion [101].

The analysis of emotion recognition confusion (Figure 7) reveals specific patterns of misidentification across emotions, regardless of expression scaling levels. Notably, *Disgust* was often mistaken for *Anger*, *Fear* for *Surprise* and *Disgust*, and *Anger* for *Disgust*. These findings align with previous research on emotion recognition challenges [32, 102, 103], suggesting that certain emotion pairs are more easily confused due to similarities in their facial expressions.

4.5.2 Uncanny Valley Effect. Interestingly, significant interaction effects were observed between *Expression Scaling* and *Emotion Category* for Facial Appeal, Facial Realism, Facial Eeriness, and Comfort (Figure 8). These interactions reveal that the impact of expression scaling on user perception varies depending on the specific emotion being conveyed.

For *Surprise*, *Joy*, and *Disgust*, increased expression scaling negatively affected perceived appeal and realism while increasing eeriness. Conversely, for *Anger*, *Fear*, and *Sadness*, higher scaling levels positively impacted these measures. Regarding Comfort, higher scaling levels significantly reduced comfort for *Joy* and *Surprise*,

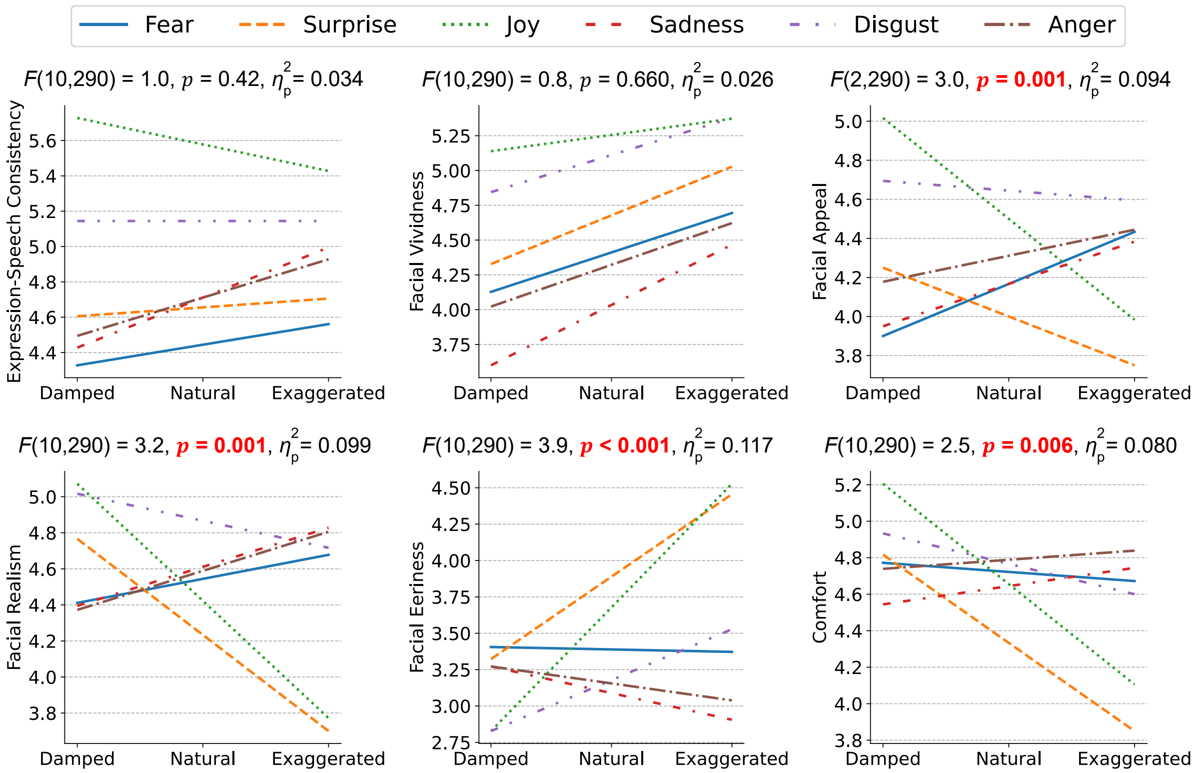


Figure 8: The interaction effect of *Emotion Category* × *Expression Scaling* on Expression-Speech Consistency (top left), Facial Vividness (top middle), Facial Appeal (top right), Facial Realism (bottom left), Facial Eeriness (bottom middle), and Comfort (bottom right).

while having minimal impact on the other four emotions. These findings highlight the complex relationship between *Expression Scaling* and *Emotion Category*, suggesting that a one-size-fits-all approach to facial expression scaling in social VR may not be optimal.

The main effects of *Expression Scaling* are illustrated in Figure 9. Facial Vividness in the *Exaggerated* condition (M = 4.94, SD = 0.18) significantly exceeds that in the *Damped* condition (M = 4.36, SD = 0.16, $p < .01$). However, Facial Realism decreases significantly in the *Exaggerated* condition (M = 4.33, SD = 0.15) compared to the *Natural* condition (M = 4.71, SD = 0.15, $p < .05$).

Moreover, Facial Eeriness in the *Exaggerated* condition (M = 3.74, SD = 0.13) is significantly higher than in both the *Damped* (M = 3.26, SD = 0.14, $p < .01$) and *Natural* conditions (M = 3.19, SD = 0.11, $p < .0001$). Concurrently, Comfort in the *Exaggerated* condition (M = 4.39, SD = 0.15) is significantly lower than in both the *Damped* (M = 4.76, SD = 0.17, $p < .05$) and *Natural* conditions (M = 4.81, SD = 0.14, $p < .001$). These results indicate that while expression exaggeration enhances vividness, it also intensifies the uncanny valley effect, reducing facial realism and user comfort.

The main effects of *Emotion Category* are illustrated in Figure 10. For Expression-Speech Consistency, *Joy* (M = 5.58, SD = 0.15) scores significantly higher than *Surprise* (M = 4.66, SD = 0.17, $p < .01$), *Fear* (M = 4.44, SD = 0.17, $p < .0001$), *Sadness* (M = 4.71, SD = 0.17, $p < .01$), and *Anger* (M = 4.71, SD = 0.18, $p < .01$). *Disgust* (M = 5.14, SD = 0.18) also rates significantly higher than *Surprise* ($p < .05$) and *Fear* ($p < .05$).

Regarding Facial Vividness, *Joy* (M = 5.26, SD = 0.21) significantly exceeds *Fear* (M = 4.41, SD = 0.19, $p < .001$), *Sadness* (M = 4.03, SD = 0.18, $p < .001$), and *Anger* (M = 4.32, SD = 0.19, $p < .05$). Similarly, *Disgust* (M = 5.11, SD = 0.19) scores significantly higher than *Fear* ($p < .001$), *Sadness* ($p < .0001$), and *Anger* ($p < .05$).

We hypothesize that the superiority of *Joy* and *Disgust* in consistency and vividness stems from their relatively stronger expression amplitudes. As shown in Table 1, these emotions exhibit higher activation levels in regions such as eyebrows, eyelid, and nose, potentially enabling more effective transmission of the actor's emotional state.

Regarding Facial Appeal and Realism, *Disgust* consistently outperforms *Surprise*. *Disgust* scores significantly higher in Appeal (M = 4.64, SD = 0.19) compared to *Surprise* (M = 4.17, SD = 0.15, $p < .05$), and similarly in Realism (*Disgust*: M = 4.87, SD = 0.19; *Surprise*: M = 4.23, SD = 0.17, $p < .05$).

The analysis of Facial Eeriness reveals interesting patterns. *Joy* (M = 3.68, SD = 0.20) elicits significantly higher eeriness than *Sadness* (M = 3.09, SD = 0.13, $p < .05$). Moreover, *Surprise* (M = 3.89, SD = 0.14) induces significantly higher eeriness compared to *Disgust* (M = 3.18, SD = 0.17, $p < .05$), *Sadness* ($p < .001$), and

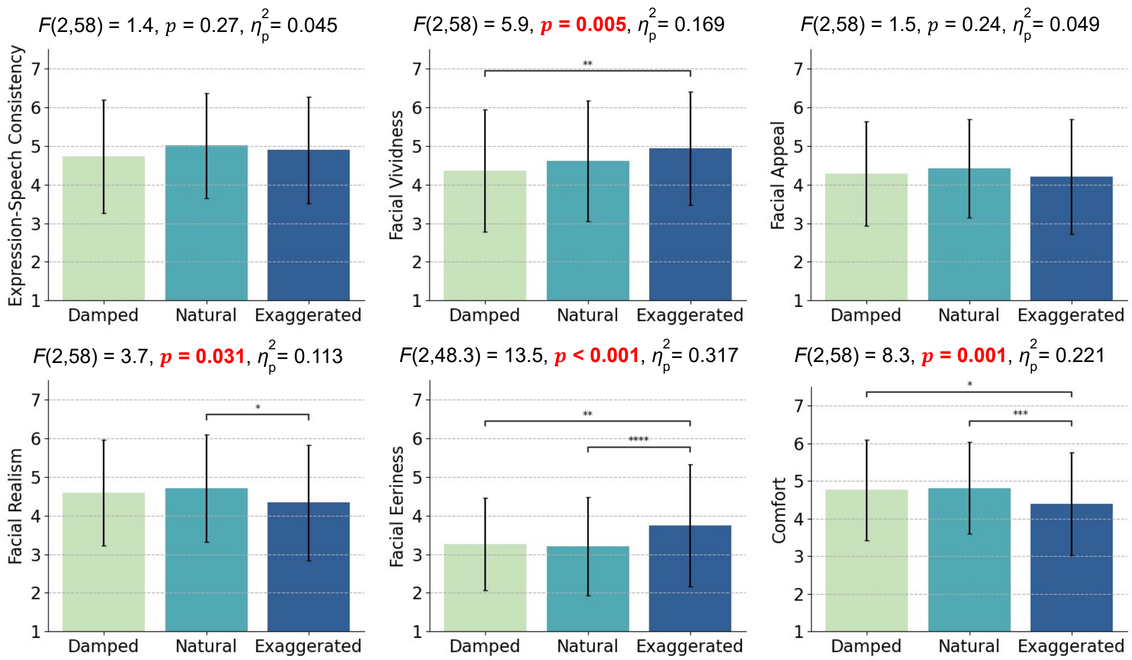


Figure 9: The main effect of *Expression Scaling* on Expression-Speech Consistency (top left), Facial Vividness (top middle), Facial Appeal (top right), Facial Realism (bottom left), Facial Eeriness (bottom middle), and Comfort (bottom right). The error bars denote standard deviation, and *, **, *, and **** denote significant differences at $p < 0.05$, $p < 0.01$, $p < 0.001$, and $p < 0.0001$ levels, respectively.**

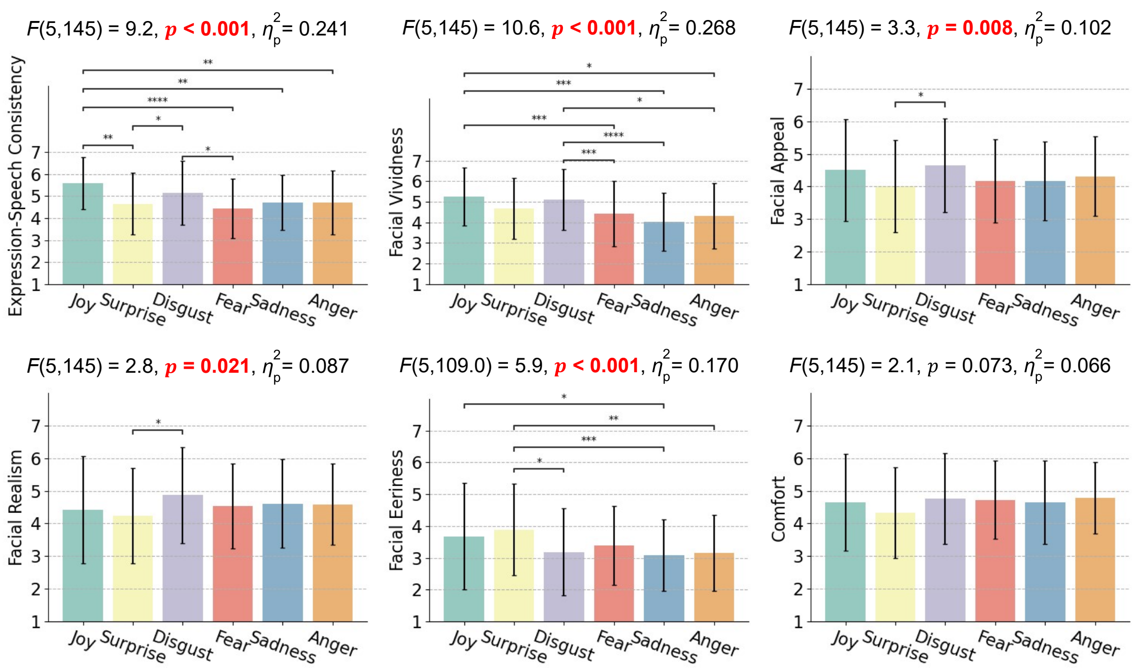


Figure 10: The main effect of *Emotion Category* on Expression-Speech Consistency (top left), Facial Vividness (top middle), Facial Appeal (top right), Facial Realism (bottom left), Facial Eeriness (bottom middle), and Comfort (bottom right). The error bars denote standard deviation, and *, **, *, and **** denote significant differences at $p < 0.05$, $p < 0.01$, $p < 0.001$, and $p < 0.0001$ levels, respectively.**

Table 5: Thematic analysis summary of open-ended answers. For different *Emotion Category*, the second column lists the non-facial and facial indicators referenced by participants for emotion recognition, with the number of mentions in parentheses (only indicators mentioned more than three times are listed in the table). The third, fourth, and fifth columns summarize the main assessments of facial expressions under *Damped*, *Natural*, and *Exaggerated* conditions, respectively. Positive assessments are highlighted in blue, and negative assessments are highlighted in red, with the number of mentions in parentheses.

Emotions	Emotional Indicators	Damped FE	Natural FE	Exaggerated FE
Joy	**Non-facial:** *speech and tone (82), laughter (40), faster pace (12), raised arms (7)* **Facial:** *smile (36), mouth corners lifting (23), raising of the cheeks (21), curved eyes (8).*	Natural smile (8), coordinated movements across different facial regions (4), look nice (4).	expressions and speech are out of sync (3), realistic curved eyes (5), believable appearance (4), vivid expression (3).	Over-exaggerated laughing (8), overly wide mouth opening (6), disproportionately high cheek raising (4).
Surprise	**Non-facial:** *speech and tone (78), sharp intake of breath (19), "Oh!" exclamations (15)* **Facial:** *widely open mouth (27), raised eyebrows (19), wide-open eyes (14).*	Natural expression (5), appropriate amplitude of facial movement (4), believable appearance (3).	Grinning too much (6), sudden changes in facial expressions (3), strange sarcastic expression (3), rich expressions (4).	Overly open jaw (11), excessively dramatic expression changes (5), disproportionately high cheek raising (4).
Disgust	**Non-facial:** *speech and tone (76), "Ugh" sounds (21), hands pushing outward (6)* **Facial:** *wrinkled nose (30), furrowed brows (20), downturned mouth corners (12), pursed lips (7), squinting eyes (5).*	Lack of facial expression other than lip movement (6), natural cheek raising (5), realistic direction of gaze (3), vivid lip curl (3).	Lacking upper face expressions (5), a lower emotional intensity than speech (4), realistic brows frowning (3).	Unnatural mouth movements (8), overly raised cheeks (5), realistic nose wrinkling (6), dynamic and expressive eyebrow frowning (4).
Fear	**Non-facial:** *speech and tone (66), trembling (20), raised shoulders (12)* **Facial:** *raised brows (17), widened eyes (9), avoidant gesture (4), averted gaze (3).*	Flat expression (11), lacking eyebrow movement (7), mouth opened too small (5), appropriate dull expression (3).	Stiff facial expression (7), limited movement with only noticeable mouth opening and closing (6), natural averted eyes (3).	Rich micro expressions (5), realistic furrowed brows (4), trembling facial muscles (3), rich micro-expressions (3).
Sadness	**Non-Facial:** *speech and tone (78), head down (40), slower pace (36), sighing (18)* **Facial:** *downward gaze (25), downturned mouth corners (18), drooping eyebrows (9).*	Looks very calm (8), almost no expression (7), face looks tense and heavy-hearted (4), expression and tone match (3).	Blank face (5), rigid facial expressions (5), a tiny smile that contradicted the emotion (5), emotional downcast eyes (4).	Vivid and authentic expression (5), realistic eyebrows drooping (5), natural lid squinting (4).
Anger	**Non-facial:** *speech and tone (74), rapid pace (36), approach posture(13), clenched fists (7)* **Facial:** *glare (29), furrowed brows (28), bared teeth (8), raising of the cheeks (5).*	Small facial expressions (12), expression does not match emotion (7), lack of upper face movement (5), stiff and tight face (4).	Flat, stiff expression with limited facial movement (8), a lower emotional intensity than speech (4).	Obvious nose wrinkling (6), naturally furrowed brows (5), expressive stare (3), vivid cheek raising (3).

Anger (M = 3.16, SD = 0.16, $p < .01$). These results suggest that expressions of *Joy* and *Surprise* trigger relatively stronger uncanny valley effects compared to other emotions.

4.6 Qualitative Results

Table 5 summarizes the thematic analysis of open-ended responses, highlighting the primary indicators participants used for emotion recognition and their assessments of damped, natural, and exaggerated facial expressions. The following sections elaborate on these findings.

4.6.1 Emotion Perception. As illustrated in the "Emotional Indicators" column of Table 5, participants reported utilizing a diverse range of modalities to perceive emotions, including speech content, tone, facial expressions, eye movements, and body posture. Notably, facial expressions constituted only 33.2% of the mentioned cues (377 mentions), while non-facial indicators accounted for the majority at 66.8%. Speech content and tone emerged as the most frequently cited indicators (454 mentions, 39.9%). Furthermore, 25 out of 30 participants regarded facial expressions as secondary in emotion judgment. This sentiment was exemplified by P17, who stated: *"I'm aware that the avatar is a synthesized virtual figure, and its expressions might not be accurate. In contrast, the voice I hear is real. Consequently, I rely more on speech to determine emotions, with facial expressions serving as a supplementary cue."* This finding underscores the multi-modal nature of emotion perception in virtual environments, where users prioritize auditory cues over visual ones in avatar-mediated interactions.

4.6.2 Uncanny Valley Effect. As illustrated in Table 5, the impact of facial expression scaling varied across six *Emotion Category*:

Joy and Surprise: expressions in *Exaggerated* condition were often perceived as unnatural, particularly in mouth movements. For *Joy*, participants noted "overexaggerated laughing" (8 mentions) and "overly wide mouth opening" (6 mentions). For example, P5 remarked on an exaggerated *Joy* clip, *"He smiled so broadly that he showed all his upper gums and even the inside of his mouth, which is very unsettling."* Similarly, for *Surprise*, "overly open jaw" (11 mentions) and "excessively dramatic expression changes" (4 mentions) were criticized. In contrast, *Natural* and *Damped* conditions were described as more "natural" and "believable," with "coordinated movements across different facial regions" (4 mentions for *Joy*) and "appropriate amplitude of facial movement" (4 mentions for *Surprise*).

Fear and Anger: expressions in *Exaggerated* condition were generally perceived as more vivid and realistic. For *Fear*, participants noted "rich micro expressions" (5 mentions) and "realistic furrowed brows" (4 mentions). For instance, P27 commented on an exaggerated *Fear* clip, *"Her brows are furrowed tightly, and her eyes convey emotions of fear and anxiety. The micro-expressions are realistic and rich."* Similarly, *Anger* showed "obvious nose wrinkling" (6 mentions) and "naturally furrowed brows" (5 mentions). Participants emphasized the importance of eyebrow movements in conveying these negative emotions, suggesting even further exaggeration (5 mentions). Conversely, *Damped* and *Natural* conditions were often described as "flat" (11 mentions for *Fear*) or "stiff" (7 mentions for *Anger*), lacking in emotional intensity.

Sadness: Opinions were mixed across scaling conditions. While *Exaggerated* expressions were seen as "vivid and authentic" (5 mentions), *Damped* expressions were described as "looking very calm" (8 mentions), and *Natural* expressions as having a "blank face" (5 mentions). Some participants perceived the lack of facial expressions in *Sadness* as realistic, as P23 noted, *"Some people express sadness with a blank, peaceful face, which I find acceptable."*

Disgust: Participants provided varied assessments across conditions. *Exaggerated* expressions enhanced realism through "dynamic and expressive eyebrow frowning" (4 mentions), but "unnatural mouth movements" (8 mentions) were criticized. In the *Damped* condition, "natural cheek raising" (5 mentions) was appreciated despite "lack of facial expression other than lip movement" (6 mentions).

These subjective evaluations explain the interaction effect observed in quantitative findings. *Joy*, *Surprise*, and *Disgust* involve large mouth movements that appear unnatural when exaggerated. Conversely, *Fear*, *Sadness*, and *Anger* primarily involve upper facial expressions, which are naturally subtle and potentially subdued by HMD pressure [123]. Exaggeration makes these subtle expressions more discernible, enhancing their perceived naturalness and realism.

5 Study 2: Validation of Region-specific Facial Expression Exaggeration Strategy

Study 1 revealed that while facial expression exaggeration generally increased the uncanny valley effect, its impact varied significantly across emotion categories. Further analysis of participants' open-ended responses suggested that this variation stems from differences in the facial regions and movement amplitudes associated with different emotions.

These findings suggest that a facial-region-specific approach to expression enhancement might be more beneficial than uniform exaggeration. Consequently, we proposed a Region-specific Facial Expression Exaggeration Strategy based on Study 1's results, aiming to amplify less pronounced facial features while maintaining the natural intensity of already expressive elements.

5.1 Region-specific Facial Expression Exaggeration Strategy

Based on findings from Study 1, our strategy applies varying degrees of exaggeration to different facial regions, tailoring enhancement to specific expression patterns and user perceptions. As shown in Figure 11, the scaling parameters for each region were determined through both quantitative analysis (Table 1) and qualitative feedback (Table 5).

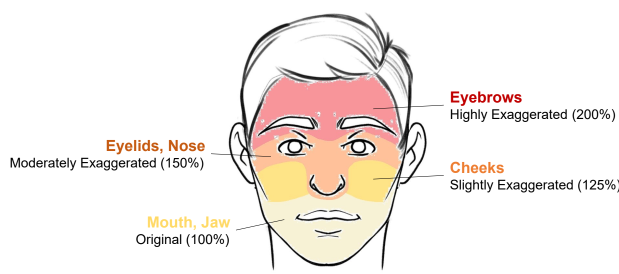


Figure 11: The illustration of Region-specific Facial Expression Exaggeration Strategy.

Eyebrows receive high exaggeration (200%) due to their crucial role in emotion conveyance [117, 126]. Study 1 revealed consistent under-expression of eyebrows in negative emotions, with participants reporting "flat" or "stiff" expressions and suggesting further exaggeration even in exaggerated conditions. This enhancement is particularly important as eyebrow movements serve as primary indicators for emotions like anger and fear.

Eyelids and nose are moderately exaggerated (150%) to address limited movement across most emotions, especially negative ones. Our data showed significantly lower activation levels in these regions (e.g., *Lid Tightener* averaging only 8.63% for fear and 3.42%

for sadness), likely due to HMD pressure effects. This scaling level compensates for under-expression while avoiding unnaturalness.

Cheeks receive slight exaggeration (125%) to balance expressiveness and naturality. While cheek movements proved crucial for disgust and anger, excessive enhancement created unnatural appearances in joy and surprise. This moderate scaling optimizes expression while avoiding uncanny valley effects.

Mouth and jaw regions maintain original activation levels, as Study 1 showed that exaggeration here significantly increased eeriness ratings, particularly for joy and surprise. These regions already demonstrated substantial natural movement (>15% average activation), making further enhancement unnecessary and potentially detrimental.

This nuanced approach addresses the region-specific challenges identified in Study 1, enhancing emotional expressiveness while maintaining natural appearance across different emotions. The following sections evaluate this strategy's effectiveness through our validation experiment.

5.2 Study Design

We conducted a comparative study using a within-subjects design, employing the same virtual environment and experimental materials as Study 1 to ensure consistency and enable direct comparison of results. The independent variable was the facial expression scaling method, with two versions: (1) our region-specific exaggeration strategy and (2) unmodified original expressions. Each participant experienced both versions of each emotional clip in randomized order. They were instructed to compare each pair of clips in terms of empathy and uncanny valley effects, allowing for direct evaluation of our strategy's impact.

5.3 Measurements

After viewing each pair, participants completed a questionnaire focusing on aspects of empathy and the uncanny valley effect, excluding emotion recognition accuracy based on Study 1 findings. They rated eight dimensions (emotional intensity, emotional contagion, expression-speech consistency, facial vividness, facial appeal, facial realism, facial eeriness, and overall comfort) using a 5-point Likert scale, responding to the statement *"The second viewed clip is better (or stronger) than the first one."* Participants were also encouraged to provide comments elaborating on their choices, offering valuable qualitative data to complement our quantitative measures.

5.4 Participants and Procedure

We recruited 20 participants (12 females, 8 males, age range: 18-29 years, M = 24.0, SD = 2.9) from the same university campus as Study 1. This sample size provided sufficient statistical power (0.8) for our within-subjects design while remaining feasible for the in-depth, VR-based study. Participants were naive to the study's objectives and had not participated in previous experiments, ensuring unbiased responses. Their VR familiarity was rated on a 5-point scale (M = 2.83, SD = 1.25), with three participants reporting experience with platforms like VRChat. The study was approved by the university's ethics committee, and participants provided written informed consent. The entire experiment lasted approximately 90 minutes, with participants receiving $20 compensation.

We randomly assigned the 10 actors to the 20 participants, with each actor's recordings being viewed by two participants. After familiarizing themselves with the experimental apparatus and virtual environment, each participant evaluated all 9 clips (in two versions) from their assigned actor, with a 30-second break between pairs to prevent fatigue. To mitigate potential order effects, we randomized the presentation order of these clips.

5.5 Analysis

We collected 180 paired comparison results (20 participants × 9 clips), retaining all data points to maintain the integrity of our dataset. Each result contains two user empathy ratings and six uncanny valley effect ratings. To ensure consistent analysis, we standardized all ratings to address the question: *"Are the region-specifically exaggerated facial expressions better (or stronger) than the unmodified original ones?"*

Our quantitative analysis followed a rigorous statistical approach. We first conducted Shapiro-Wilk tests to assess the normality of distribution for each measure. For normally distributed data, we employed One-Sample T-tests to compare the mean rating against the neutral point of 3 ("neither agree nor disagree") on our 5-point Likert scale. For non-normally distributed data, we applied the non-parametric Mann-Whitney U test as an alternative, still comparing against the neutral point of 3. This method allowed us to determine whether participants showed a statistically significant preference for either the region-specifically exaggerated or unmodified original expressions across various measures.

5.6 Results

Figure 12 illustrates participants' ratings of our region-specific exaggeration strategy across different emotional categories and dimensions. The strategy's effectiveness varied among emotions:

Disgust and Anger: The strategy was highly effective, showing significant improvements across all measured dimensions (all $p < .05$ for *Disgust*, $p < .01$ for *Anger*). For *Disgust*, participants noted enhanced naturalness and speech alignment. For *Anger*, eyebrow exaggeration was particularly impactful, with participants reporting more realistic and intense emotional conveyance. For example, P6 commented: *"The frown and wide eyes in the second (region-specifically exaggerated) clip is more real, like staring at me. I feel his strong anger."*

Fear: The strategy showed significant improvements in all dimensions except Facial Eeriness. Participants highlighted that exaggerated upper facial features (frowning, glaring, squinting) contributed to stronger and more realistic emotional expression.

Sadness: Only Facial Realism showed significant improvement ($p < .05$). Participants often struggled to detect differences between original and exaggerated versions, possibly due to the naturally subtle expressions associated with sadness.

Joy and Surprise: No statistically significant differences were observed. While some participants appreciated enhanced upper facial expressions, others noted potential issues with exaggerated eyebrow and eyelid movements (e.g., pronounced eyebrow movements in joyful expressions created an incongruent sense of frowning), suggesting a delicate balance is needed for these emotions.

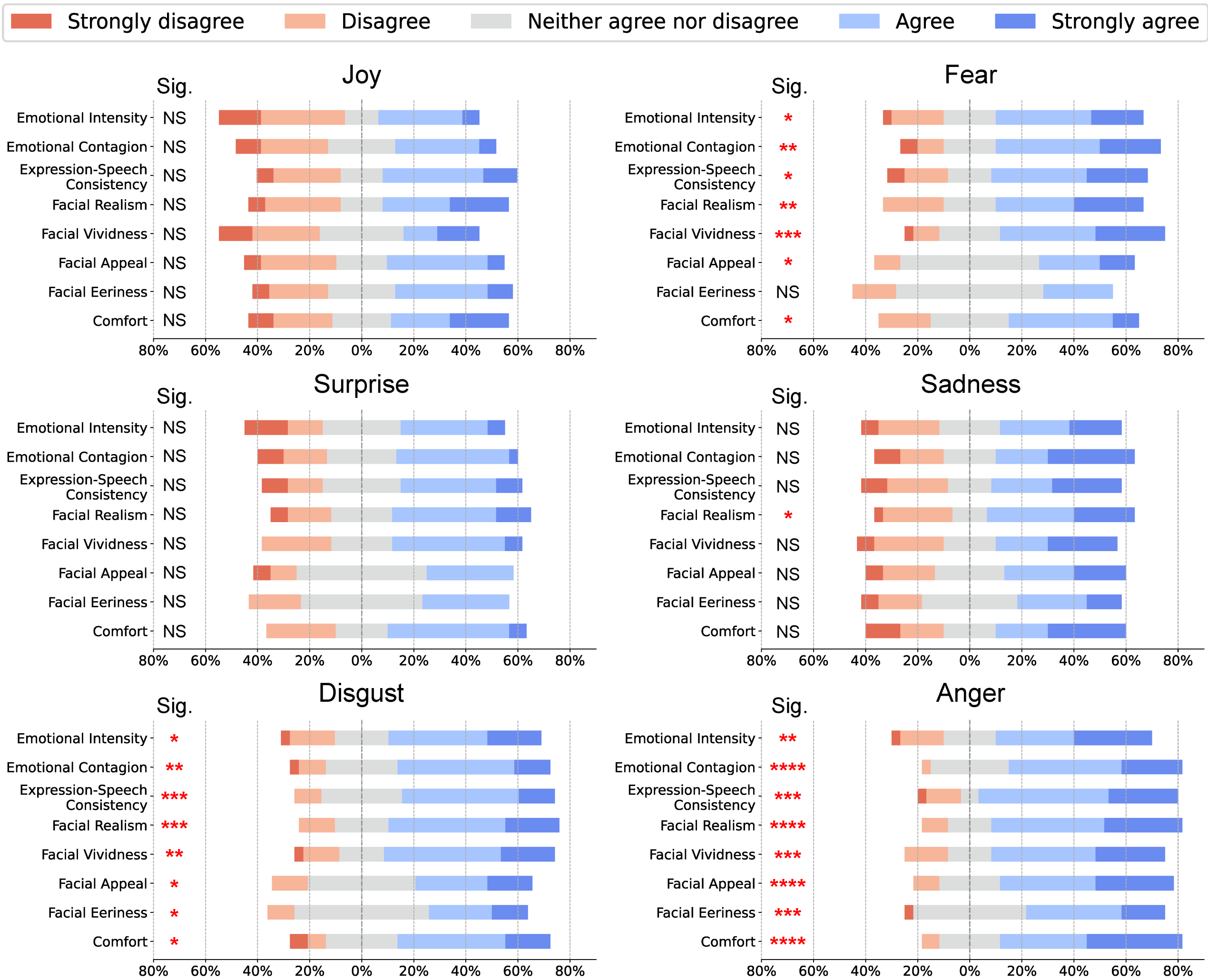


Figure 12: Participants' ratings for the question: "Are the perception of region-specifically exaggerated facial expressions better (or stronger) than the unmodified original ones?" Results are categorized by emotional categories and various dimensions of user empathy and uncanny valley effect. Statistical tests were conducted to compare the mean rating against the neutral point. Significance levels are indicated as follows: * p < .05, ** p < .01, *** p < .001, **** p < .0001. NS indicates no statistically significant difference.

In summary, our region-specific exaggeration strategy proved most effective for negative emotions (*Disgust*, *Anger*, *Fear*), enhancing both emotional conveyance and perceived realism. It had minimal impact on *Joy* and *Surprise*, and limited effect on *Sadness*. These results suggest that the strategy's effectiveness is emotion-dependent, with greater benefits for emotions typically expressed through subtle upper facial movements.

6 Discussion

This study investigated how facial expression scaling affects user empathy and the uncanny valley effect in social VR environments, addressing two key questions: *How does expression scaling influence emotional perception and the uncanny valley effect? (RQ1)*, and *how can we leverage scaling to enhance emotional communication while mitigating potential uncanny valley effects? (RQ2)*

Our findings challenge traditional assumptions about emotion recognition in virtual environments. While expression scaling did not significantly affect recognition accuracy, we discovered a novel interaction effect between scaling and emotion types on the uncanny valley effect. This interaction demonstrates that scaling impacts vary substantially across different emotions, leading to our development of a region-specific expression exaggeration strategy. This approach effectively enhanced emotional conveyance for negative emotions while mitigating uncanny valley effects.

These insights advance both theoretical understanding of virtual emotional communication and practical avatar design. The following sections examine our findings in detail, exploring their implications for social VR development and proposing specific design considerations for future applications.

6.1 The Effect of Facial Expression Scaling on User Empathy (RQ1)

Our research provides novel insights into the complex relationship between facial expression scaling and user empathy in social VR. Study 1 findings (Figure 5) reveal that facial expression scaling does not significantly affect recognition accuracy of emotional category or intensity in avatar-mediated interactions. This result diverges from traditional emotion recognition paradigms that emphasize the primacy of facial cues [8, 19, 46, 60].

Further investigation (Table 5) shows participants' preference for non-facial cues, particularly speech content and tone, over facial expressions in avatar-mediated communication. This shift from visual to auditory cues suggests a fundamental difference in emotional information processing between virtual and face-to-face interactions. Our findings extend previous research on trust in virtual environments [97], indicating that users prioritize authentic vocal cues over synthesized visual expressions in social VR. These results underscore the need for multi-modal virtual avatars that integrate facial expressions with voice tone and body language, aligning with users' natural tendency to rely on multiple cues for emotion perception in virtual environments.

Study 2 further elucidates the nuanced nature of user empathy in VR, highlighting the critical role of upper facial expressions in conveying negative emotions. As shown in Figure 12, our region-specific exaggeration strategy significantly improved the conveyance of fear, disgust, and anger by enhancing upper facial movements. These findings corroborate that emotion perception relies on different facial action units [107] and can be organized along the upper-lower face axis [71]. Upper facial expressions are particularly important in perceiving negative emotions [117, 126], underscoring a crucial limitation of current VR systems: HMD pressure potentially suppressing vital expressive movements [123]. Our research demonstrates how region-specific exaggeration can compensate for these physical limitations, providing valuable insights for designing more emotionally expressive avatar systems.

By examining user empathy in immersive, dynamic social VR settings, our research bridges a significant gap in the literature, moving beyond the limitations of static image and 2D screen-based avatar studies [16, 34, 47, 54, 74, 93]. This approach provides a more ecologically valid understanding of emotional communication in virtual environments, paving the way for more nuanced and effective avatar design in social VR applications.

6.2 The Effect of Facial Expression Scaling on Uncanny Valley Effect (RQ1)

Our research offers a novel perspective on the uncanny valley effect in social VR, grounded in recent neuroscientific theories of predictive brain function. These theories suggest that the brain continuously anticipates sensory input based on past experiences, with prediction signals significantly influencing perception [12]. We extend this concept to avatar perception, proposing that the uncanny valley effect arises when facial expressions deviate from users' anticipations, which are formed based on contextual and multimodal cues.

Through quantitative analysis (Table 4) and qualitative assessment (Table 5), we uncovered that varying facial action unit movement patterns across emotions result in different impacts of expression scaling on users' anticipations. For joy, surprise, and disgust, which involve pronounced mouth movements, exaggerated expressions often exceeds users' natural anticipations, triggering the uncanny valley effect. Conversely, for sadness, disgust, and anger, upper facial expressions are typically subdued in VR due to HMD pressure [123]. Without scaling or in damped conditions, expression intensity falls below users' anticipations, leading to lower perceived realness and naturalness. In these cases, exaggeration can actually enhance recognizability and reduce the uncanny valley effect.

These findings reveal a crucial interaction effect between expression scaling and emotion types, advancing our understanding of the uncanny valley effect beyond previous research that primarily focused on the general risks of expression exaggeration [47, 74]. Our work demonstrates that the impact of expression scaling is not uniform but highly dependent on the specific emotion being conveyed, challenging the notion that consistently scaled expressions are universally preferable [35, 87]. This nuanced understanding suggests that optimal expression scaling should be region-specific and emotion-adaptive, aligning with users' anticipations.

6.3 Leveraging Region-specific Facial Expression Exaggeration to Enhance Emotional Communication while Mitigating Uncanny Valley Effects (RQ2)

Our research introduces a novel approach to enhancing emotional communication in social VR through region-specific facial expression scaling (illustrated in Figure 11), extending beyond previous efforts that primarily focused on enhancing positive emotions [87] or selecting prioritized facial expressions [51]. As demonstrated in Study 2 (Figure 12), this strategy not only increases perceived emotional intensity and contagion for negative emotions like anger, fear, and disgust but also mitigates the uncanny valley effect, addressing a critical challenge in avatar-mediated interactions. By tailoring expression scaling to specific facial areas based on empirical findings, we've shown it's possible to enhance emotional expressiveness while maintaining perceived naturalness, crucial for creating engaging and empathetic virtual social experiences. These findings not only contributes to the theoretical understanding of the emotion conveyance and uncanny valley effect but also provides practical guidelines for creating more effective and authentic avatar-mediated communications in social VR.

While our region-specific strategy represents a significant advancement in avatar expression systems, its varying effectiveness across emotions highlights the need for more nuanced approaches. As illustrated in Section 5.6, for joy and surprise, although the strategy doesn't significantly increase the uncanny valley effect, user feedback indicates that exaggerated eyebrow and eyelid movements can create artifacts in positive emotional expressions, potentially negatively impacting user experience. Additionally, the strategy's limited effect on sadness, an emotion with inherently low expression amplitude, suggests that current scaling amplitude designs may not be "sweet spots" for all emotions. These findings indicate that optimal enhancement strategies should be both region-specific and emotion-dependent.

Looking forward, our research lays the groundwork for integrating region-specific expression scaling with advanced emotion recognition technologies, potentially leading to dynamic, context-aware avatar expression systems. Such systems could revolutionize emotional communication in virtual environments, making them more natural and emotionally resonant. Our findings also contribute to the broader field of affective computing in VR, providing insights into how different emotions activate specific facial regions and the appropriate scaling for these movements. This knowledge can inform cutting-edge work in computational animation and machine learning approaches to avatar expression generation [4, 90, 91, 111].

6.4 Generalizability of Our Findings

Our study design incorporated measures to control for individual differences in emotional expression and perception, including gender [6], personality traits [67], and transient mood [125]. This involved recruiting a substantial participant pool (10 actors, 50 participants in Study 1, and 20 in Study 2) with balanced gender ratios. However, several important limitations affect the generalizability of our findings.

The primary limitation stems from demographic homogeneity. Our participant pool consisted predominantly of Chinese individuals aged 20-30 years, which, while enabling controlled comparisons, constrains cross-cultural generalizability. Although basic emotion categories are universal [28], significant cultural variations exist in emotional expression and perception [50]. For instance, Westerners and Easterners employ different facial movement patterns to express basic emotions, with East Asian individuals typically displaying more conservative expressions due to collectivist cultural influences [105]. Cultural differences also manifest in judgments of emotional intensity [28], suggesting our region-specific scaling strategy may require cultural adaptation.

Age-related factors present another limitation. Our focus on young adults overlooks important age-related variations in emotion perception documented in previous research. Studies have shown that older adults generally demonstrate lower emotion-perception accuracy than younger individuals [84] and tend to interpret emotional stimuli as more positive and arousing [85]. These age-related differences suggest that the optimal scaling degrees we identified for different facial regions may need calibration across age groups.

A third limitation concerns avatar design choices. Our findings are based on cartoon-style avatars and may not directly translate to photo-realistic applications. Research indicates that increased fidelity and photorealism can enhance avatar attractiveness and humanlikeness while reducing eeriness [10, 39], suggesting that avatar appearance fidelity might offer an alternative or complementary approach to mitigating the uncanny valley effect.

These limitations highlight three critical directions for future research: validating and adjusting facial region scaling parameters across different cultural groups, investigating age-related variations in response to facial expression scaling, and examining the interaction between expression scaling and avatar fidelity levels. Such extensions would enhance the robustness and applicability of our findings across diverse user populations and virtual reality applications.

6.5 Ethical Considerations

The implementation of facial expression scaling in social VR raises important ethical considerations that require careful attention. We discuss key challenges and propose guidelines for responsible deployment while maximizing the benefits of enhanced emotional communication.

Privacy and data security represent paramount concerns. Social VR platforms implementing facial expression scaling must establish robust protocols for data handling, including secure storage, encrypted transmission, and clear retention policies. This is particularly crucial as facial data could be exploited for identity-based attacks [109]. Platforms must obtain explicit user consent before recording and processing facial expressions, providing transparent explanations of data usage and protection measures.

Equally important is transparency in technical implementation. Platforms should clearly communicate both the purpose and underlying principles of facial expression scaling to users. This communication should emphasize that the technology enhances expression fidelity rather than manipulates emotional content, helping users understand that scaling serves to improve the authenticity of emotional conveyance rather than alter its fundamental nature.

Our research suggests that facial expression scaling could actually enhance privacy when properly implemented. Platforms could integrate scaling options within face modulation systems, allowing users to intentionally modify their expressions to preserve anonymity while maintaining effective emotional communication [104]. This creates a balanced framework where users can benefit from enhanced expressiveness while retaining control over their digital presence.

7 Limitations and Future Work

While our study advances understanding of facial expression scaling in social VR, several methodological constraints and limitations warrant discussion and suggest future research directions.

Our primary methodological trade-off involves the use of prerecorded, scripted interactions instead of real-time communication. While this controlled approach was necessary to isolate expression scaling effects and ensure consistent emotional content across conditions, it sacrifices ecological validity. Real-time social VR interactions involve dynamic emotional exchanges, mutual adaptation, and spontaneous expressions that our current methodology cannot capture [31]. This limitation affects our understanding of how expression scaling might function in truly interactive scenarios, particularly regarding temporal dynamics and reciprocal emotional influences. Future studies should progressively incorporate more naturalistic elements while maintaining experimental control, perhaps through semi-structured interactions or controlled yet interactive scenarios.

The scope of our interaction contexts presents another significant limitation. By focusing on dyadic interactions in static, indoor environments, we established a clear baseline but overlooked important social and environmental factors. Group dynamics fundamentally alter emotion transmission and facial mimicry patterns [7, 89], while body posture [2, 80] and environmental context [98, 99] significantly influence expression perception. Future work should systematically examine how these contextual factors interact with facial expression scaling, particularly in multi-user scenarios and diverse virtual environments.

While we demonstrated our region-specific scaling strategy's effectiveness, optimal scaling parameters require more rigorous investigation. Future studies should move beyond comparing discrete scaling conditions to employ comprehensive quantitative analyses investigating the continuous relationships between scaling ratios and emotional perception. This approach would not only help identify "sweet spots" for different facial regions and emotions but could also reveal important interaction effects between regions that our current discrete-condition approach may have missed. These insights would enable development of more sophisticated, context-aware scaling algorithms that better adapt to different emotional expressions and user preferences.

Looking forward, several promising directions emerge. Longitudinal studies should examine user adaptation to scaled expressions and their impact on critical social VR outcomes such as trust-building and collaboration efficacy. Development of context-aware scaling systems that dynamically adjust to interaction scenarios could further advance the field. These extensions would help bridge the gap between controlled experimental findings and practical applications in dynamic social VR environments.

8 Conclusion

This study provides a comprehensive investigation into the effects of facial expression scaling on user empathy and the uncanny valley effect in social VR. Through a rigorous two-part experimental design, we uncovered nuanced relationships between expression exaggeration, emotion categories, and user experience. Our findings challenge the assumption that uniform expression enhancement universally improves emotional communication in virtual environments. Key contributions include demonstrating the robustness of multimodal emotion perception in VR, revealing significant interaction effects between expression scaling and emotion categories on the uncanny valley effect, and developing a region-specific facial expression exaggeration strategy that effectively enhances emotional conveyance while mitigating uncanny valley effects. These insights not only extend our understanding of emotional communication in virtual environments but also offer practical guidelines for creating more expressive and empathetic avatar systems, laying the groundwork for developing more nuanced, effective, and user-centric approaches to virtual interaction design across various domains of remote communication.

Acknowledgments

This work was supported by the Natural Science Foundation of China under Grant No. 62132010, 62472243, and 2022YFB3105201. We would like to express our gratitude to Lingling Wang, a 3D artist from the Academy of Arts & Design at Tsinghua University, for her invaluable assistance with facial expression animation. We are also grateful to Peter Yao for his support. Our thanks extend to all the actors and participants who contributed to our experiments. Finally, we would like to thank all anonymous reviewers and Area Chairs for their constructive feedback and suggestions.

References

[1] Sahar Aseeri and Victoria Interrante. 2021. The influence of avatar representation on interpersonal communication in virtual social environments. *IEEE transactions on visualization and computer graphics* 27, 5 (2021), 2608–2617.

[2] Hillel Aviezer, Ran R Hassin, Jennifer Ryan, Cheryl Grady, Josh Susskind, Adam Anderson, Morris Moscovitch, and Shlomo Bentin. 2008. Angry, disgusted, or afraid? Studies on the malleability of emotion perception. *Psychological science* 19, 7 (2008), 724–732.

[3] Jo-Anne Bachorowski. 1999. Vocal expression and perception of emotion. *Current directions in psychological science* 8, 2 (1999), 53–57.

[4] Zechen Bai, Peng Chen, Xiaolan Peng, Lu Liu, Naiming Yao, and Hui Chen. 2024. Bring Your Own Character: A Holistic Solution for Automatic Facial Animation Generation of Customized Characters. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. IEEE, 429–438.

[5] Jakki O. Bailey and Isabella Schloss. 2023. "Awesomely Freaky!" The Impact of Type on Children's Social-Emotional Perceptions of Virtual Reality Characters. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–10. https://doi.org/10.1145/3544548.3581501

[6] Cezary Biele and Anna Grabowska. 2006. Sex differences in perception of emotion intensity in dynamic and static facial expressions. *Experimental brain research* 171 (2006), 1–6.

[7] Lauren Buck, Gareth W. Young, and Rachel McDonnell. 2023. Avatar Customization, Personality, and the Perception of Work Group Inclusion in Immersive Virtual Reality. In *Computer Supported Cooperative Work and Social Computing*. ACM, Minneapolis MN USA, 27–32. https://doi.org/10.1145/3584931.3606992

[8] Manuel G Calvo, Aida Gutiérrez-García, Andrés Fernández-Martín, and Lauri Nummenmaa. 2014. Recognition of facial expressions of emotion is related to their frequency in everyday life. *Journal of Nonverbal Behavior* 38 (2014), 549–567.

[9] Qiongdan Cao, Hui Yu, Paul Charisse, Si Qiao, and Brett Stevens. 2023. Is High-Fidelity Important for Human-like Virtual Avatars in Human Computer Interactions? *International Journal of Network Dynamics and Intelligence* (March 2023), 15–23. https://doi.org/10.53941/ijndi0201008

[10] Qiongdan Cao, Hui Yu, Paul Charisse, Si Qiao, and Brett Stevens. 2023. Is high-fidelity important for human-like virtual avatars in human computer interactions? *International Journal of Network Dynamics and Intelligence* (2023), 15–23.

[11] Ho-Seung Cha, Seong-Jun Choi, and Chang-Hwan Im. 2020. Real-time recognition of facial expressions using facial electromyograms recorded around the eyes for social virtual reality applications. *IEEE Access* 8 (2020), 62065–62075.

[12] Lorena Chanes, Jolie Baumann Wormwood, Nicole Betz, and Lisa Feldman Barrett. 2018. Facial expression predictions as drivers of social perception. *Journal of personality and social psychology* 114, 3 (2018), 380.

[13] Zubin Choudhary, Nahal Norouzi, Austin Erickson, Ryan Schubert, Gerd Bruder, and Gregory F. Welch. 2023. Exploring the Social Influence of Virtual Humans Unintentionally Conveying Conflicting Emotions. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. IEEE, Shanghai, China, 571–580. https://doi.org/10.1109/VR55154.2023.00072

[14] Theo Combe, Rebecca Fribourg, Lucas Detto, and Jean-Marie Normand. 2024. Exploring the Influence of Virtual Avatar Heads in Mixed Reality on Social Presence, Performance and User Experience in Collaborative Tasks. *IEEE Transactions on Visualization and Computer Graphics* (2024), 1–11. https://doi.org/10.1109/TVCG.2024.3372051

[15] Juliet M Corbin and Anselm Strauss. 1990. Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative sociology* 13, 1 (1990), 3–21.

[16] Amy Dawel, Elizabeth J Miller, Annabel Horsburgh, and Patrice Ford. 2022. A systematic survey of face stimuli used in psychological research 2000–2020. *Behavior Research Methods* 54, 4 (2022), 1889–1901.

[17] Juan Del Aguila, Luz M. González-Gualda, María Angeles Játiva, Patricia Fernández-Sotos, Antonio Fernández-Caballero, and Arturo S. García. 2021. How Interpersonal Distance Between Avatar and Human Influences Facial Affect Recognition in Immersive Virtual Reality. *Frontiers in Psychology* 12 (July 2021), 675515. https://doi.org/10.3389/fpsyg.2021.675515

[18] Attilio Della Greca, Amaro Ilaria, Cesare Tucci, Nicola Frugieri, and Genoveffa Tortora. 2024. A user study on the relationship between empathy and facial-based emotion simulation in Virtual Reality. In *Proceedings of the 2024 International Conference on Advanced Visual Interfaces*. 1–9.

[19] Artemisa R Dores, Fernando Barbosa, Cristina Queirós, Irene P Carvalho, and Mark D Griffiths. 2020. Recognizing emotions through facial expressions: A largescale experimental study. *International journal of environmental research*

and public health 17, 20 (2020), 7420.
[20] Shichuan Du, Yong Tao, and Aleix M Martinez. 2014. Compound facial expressions of emotion. *Proceedings of the national academy of sciences* 111, 15 (2014), E1454–E1462.
[21] Changming Duan and Clara E Hill. 1996. The current state of empathy research. *Journal of counseling psychology* 43, 3 (1996), 261.
[22] Charlotte Dubosc, Geoffrey Gorisse, Olivier Christmann, Sylvain Fleury, Killian Poinsot, and Simon Richir. 2021. Impact of Avatar Facial Anthropomorphism on Body Ownership, Attractiveness and Social Presence in Collaborative Tasks in Immersive Virtual Environments. *Computers & Graphics* 101 (Dec. 2021), 82–92. https://doi.org/10.1016/j.cag.2021.08.011
[23] Miriam Dyck, Maren Winbeck, Susanne Leiberg, Yuhan Chen, Rurben C Gur, and Klaus Mathiak. 2008. Recognition profile of emotions in natural and virtual faces. *PloS one* 3, 11 (2008), e3628.
[24] Elena Dzardanova, Vasiliki Nikolakopoulou, Vlasios Kasapakis, Spyros Vosinakis, Ioannis Xenakis, and Damianos Gavalas. 2024. Exploring the Impact of Non-verbal Cues on User Experience in Immersive Virtual Reality. *Computer Animation and Virtual Worlds* 35, 1 (Jan. 2024), e2224. https://doi.org/10.1002/cav.2224
[25] Paul Ekman. 1982. Methods for measuring facial action. *Handbook of methods in nonverbal behavior research* (1982), 45–90.
[26] Paul Ekman. 1992. An argument for basic emotions. *Cognition & emotion* 6, 3-4 (1992), 169–200.
[27] Paul Ekman and Wallace V Friesen. 1971. Constants across cultures in the face and emotion. *Journal of personality and social psychology* 17, 2 (1971), 124.
[28] Paul Ekman, Wallace V Friesen, Maureen O'sullivan, Anthony Chan, Irene Diacoyanni-Tarlatzis, Karl Heider, Rainer Krause, William Ayhan LeCompte, Tom Pitcairn, Pio E Ricci-Bitti, et al. 1987. Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of personality and social psychology* 53, 4 (1987), 712.
[29] Hillary Anger Elfenbein and Nalini Ambady. 2003. When familiarity breeds accuracy: cultural exposure and facial emotion recognition. *Journal of personality and social psychology* 85, 2 (2003), 276.
[30] Lisa A Elkin, Matthew Kay, James J Higgins, and Jacob O Wobbrock. 2021. An aligned rank transform procedure for multifactor contrast tests. In *The 34th annual ACM symposium on user interface software and technology*. 754–768.
[31] Guo Freeman and Divine Maloney. 2021. Body, Avatar, and Me: The Presentation and Perception of Self in Social Virtual Reality. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (Jan. 2021), 1–27. https://doi.org/10.1145/3432938
[32] C.N.W. Geraets, S. Klein Tuente, B.P. Lestestuiver, M. Van Beilen, S.A. Nijman, J.B.C. Marsman, and W. Veling. 2021. Virtual Reality Facial Emotion Recognition in Social Environments: An Eye-Tracking Study. *Internet Interventions* 25 (Sept. 2021), 100432. https://doi.org/10.1016/j.invent.2021.100432
[33] Mar Gonzalez-Franco, Anthony Steed, Steve Hoogendyk, and Eyal Ofek. 2020. Using Facial Animation to Increase the Enfacement Illusion and Avatar Self-Identification. *IEEE Transactions on Visualization and Computer Graphics* 26, 5 (May 2020), 2023–2029. https://doi.org/10.1109/TVCG.2020.2973075
[34] Jonathon D. Hart, Thammathip Piumsomboon, Louise Lawrence, Gun A. Lee, Ross T. Smith, and Mark Billinghurst. 2018. Emotion Sharing and Augmentation in Cooperative Virtual Reality Games. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts*. ACM, Melbourne VIC Australia, 453–460. https://doi.org/10.1145/3270316.3271543
[35] Jonathon Derek Hart, Thammathip Piumsomboon, Gun A. Lee, Ross T. Smith, and Mark Billinghurst. 2021. Manipulating Avatars for Enhanced Communication in Extended Reality. In *2021 IEEE International Conference on Intelligent Reality (ICIR)*. IEEE, Piscataway, NJ, USA, 9–16. https://doi.org/10.1109/ICIR51845.2021.00011
[36] Jakob Hartbrich, Florian Weidner, Christian Kunert, Alexander Raake, Wolfgang Broll, and Stephanie Arévalo Arboleda. 2023. Eye and Face Tracking in VR: Avatar Embodiment and Enfacement with Realistic and Cartoon Avatars. In *Proceedings of the 22nd International Conference on Mobile and Ubiquitous Multimedia*. ACM, Vienna Austria, 270–278. https://doi.org/10.1145/3626705.3627793
[37] Viviane Herdel, Anastasia Kuzminykh, Andrea Hildebrandt, and Jessica R. Cauchard. 2021. Drone in Love: Emotional Perception of Facial Expressions on Flying Robots. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–20. https://doi.org/10.1145/3411764.3445495
[38] Fernanda Herrera, Soo Youn Oh, and Jeremy N Bailenson. 2020. Effect of behavioral realism on social interactions inside collaborative virtual environments. *Presence* 27, 2 (2020), 163–182.
[39] Darragh Higgins, Donal Egan, Rebecca Fribourg, Benjamin Cowan, and Rachel McDonnell. 2021. Ascending from the valley: Can state-of-the-art photorealism avoid the uncanny?. In *ACM Symposium on Applied Perception 2021*. 1–5.
[40] Darragh Higgins, Yilin Zhan, Benjamin R. Cowan, and Rachel McDonnell. 2023. Investigating the Effect of Visual Realism on Empathic Responses to Emotionally Expressive Virtual Humans. In *ACM Symposium on Applied Perception 2023*. ACM, Los Angeles CA USA, 1–7. https://doi.org/10.1145/3605495.3605799
[41] Darragh Higgins, Katja Zibrek, Joao Cabral, Donal Egan, and Rachel McDonnell. 2022. Sympathy for the digital: Influence of synthetic voice on affinity, social presence and empathy for photorealistic virtual humans. *Computers & Graphics* 104 (2022), 116–128.
[42] Chin-Chang Ho and Karl F. MacDorman. 2017. Measuring the Uncanny Valley Effect: Refinements to Indices for Perceived Humanness, Attractiveness, and Eeriness. *International Journal of Social Robotics* 9, 1 (Jan. 2017), 129–139. https://doi.org/10.1007/s12369-016-0380-9
[43] Ruud Hortensius, Felix Hekele, and Emily S Cross. 2018. The perception of emotion in artificial agents. *IEEE Transactions on Cognitive and Developmental Systems* 10, 4 (2018), 852–864.
[44] Natalie Hube, Oliver Lenz, Lars Engeln, Rainer Groh, and Michael Sedlmair. 2020. Comparing Methods for Mapping Facial Expressions to Enhance Immersive Collaboration with Signs of Emotion. In *2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, Recife, Brazil, 30–35. https://doi.org/10.1109/ISMAR-Adjunct51615.2020.00023
[45] Natalie Hube, Kresimir Vidackovic, and Michael Sedlmair. 2022. Using Expressive Avatars to Increase Emotion Recognition: A Pilot Study. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. ACM, New Orleans LA USA, 1–7. https://doi.org/10.1145/3491101.3519822
[46] Stephan Huber and Natalie Rathß. 2023. Empathic Accuracy and Mental Effort During Remote Assessments of Emotions. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–13. https://doi.org/10.1145/3544548.3580824
[47] Jennifer Hyde, Elizabeth J. Carter, Sara Kiesler, and Jessica K. Hodgins. 2013. Perceptual Effects of Damped and Exaggerated Facial Motion in Animated Characters. In *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*. IEEE, Shanghai, China, 1–6. https://doi.org/10.1109/FG.2013.6553775
[48] Jennifer Hyde, Elizabeth J. Carter, Sara Kiesler, and Jessica K. Hodgins. 2015. Using an Interactive Avatar's Facial Expressiveness to Increase Persuasiveness and Socialness. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, Seoul Republic of Korea, 1719–1728. https://doi.org/10.1145/2702123.2702465
[49] CARROLL E Izard. 1978. Emotions as motivations: an evolutionary-developmental perspective.. In *Nebraska symposium on motivation*. University of Nebraska Press.
[50] Rachael E Jack, Oliver GB Garrod, Hui Yu, Roberto Caldara, and Philippe G Schyns. 2012. Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences* 109, 19 (2012), 7241–7244.
[51] Seoyoung Kang, Hail Song, Boram Yoon, Kangsoo Kim, and Woontack Woo. 2024. The Influence of Emotion-based Prioritized Facial Expressions on Social Presence in Avatar-mediated Remote Communication. In *2024 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 1147–1156.
[52] Seoyoung Kang, Boram Yoon, Bowon Kim, and Woontack Woo. 2022. Effects of Avatar Face Level of Detail Control on Social Presence in Augmented Reality Remote Collaboration. In *2022 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 763–767.
[53] Jari Kätsyri, Klaus Förger, Meeri Mäkäräinen, and Tapio Takala. 2015. A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness. *Frontiers in psychology* 6 (2015), 390.
[54] Lorena C Kegel, Peter Brugger, Sascha Frühholz, Thomas Grunwald, Peter Hilfiker, Oona Kohnen, Miriam L Loertscher, Dieter Mersch, Anton Rey, Teresa Sollfrank, et al. 2020. Dynamic human and avatar facial expressions elicit differential brain responses. *Social cognitive and affective neuroscience* 15, 3 (2020), 303–317.
[55] Simon Kimmel, Frederike Jung, Andrii Matviienko, Wilko Heuten, and Susanne Boll. 2023. Let's Face It: Influence of Facial Expressions on Social Presence in Collaborative Virtual Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–16. https://doi.org/10.1145/3544548.3580707
[56] Julian Kiverstein. 2015. Empathy and the responsiveness to social affordances. *Consciousness and Cognition* 36 (2015), 532–542.
[57] Christian G Kohler, Travis Turner, Neal M Stolar, Warren B Bilker, Colleen M Brensinger, Raquel E Gur, and Ruben C Gur. 2004. Differences in facial expressions of four universal emotions. *Psychiatry research* 128, 3 (2004), 235–244.
[58] Elena Kokkinara and Rachel McDonnell. 2015. Animation Realism Affects Perceived Character Appeal of a Self-Virtual Face. In *Proceedings of the 8th ACM SIGGRAPH Conference on Motion in Games*. ACM, Paris France, 221–226. https://doi.org/10.1145/2822013.2822035
[59] Mariska E Kret, Karin Roelofs, Jeroen J Stekelenburg, and Beatrice de Gelder. 2013. Emotional signals from faces, bodies and scenes influence observers' face expressions, fixations and pupil-size. *Frontiers in human neuroscience* 7 (2013), 810.
[60] Eva G Krumhuber, Lina I Skora, Harold CH Hill, and Karen Lander. 2023. The role of facial movements in emotion recognition. *Nature Reviews Psychology* 2,

5 (2023), 283–296.
[61] Peter Kullmann, Timo Menzel, Mario Botsch, and Marc Erich Latoschik. 2023. An Evaluation of Other-Avatar Facial Animation Methods for Social VR. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–7. https://doi.org/10.1145/3544549.3585617
[62] Dilshani Kumarapeli, Sungchul Jung, and Robert W. Lindeman. 2022. Emotional Avatars: Effect of Uncanniness in Identifying Emotions Using Avatar Expressions. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, Christchurch, New Zealand, 650–651. https://doi.org/10.1109/VRW55335.2022.00176
[63] Marc Erich Latoschik, Daniel Roth, Dominik Gall, Jascha Achenbach, Thomas Waltemate, and Mario Botsch. 2017. The effect of avatar realism in immersive social virtual realities. In *Proceedings of the 23rd ACM symposium on virtual reality software and technology*. 1–10.
[64] Petri Laukka, Patrik Juslin, and Roberto Bresin. 2005. A dimensional approach to vocal expression of emotion. *Cognition & Emotion* 19, 5 (2005), 633–653.
[65] Ching-Long Lee, Wen Pei, Yu-Cheng Lin, Anders Granmo, and Kang-Hung Liu. 2023. Emotion detection based on pupil variation. In *Healthcare*, Vol. 11. MDPI, 322.
[66] Sueyoon Lee, Abdallah El Ali, Maarten Wijntjes, and Pablo Cesar. 2022. Understanding and Designing Avatar Biosignal Visualizations for Social Virtual Reality Entertainment. In *CHI Conference on Human Factors in Computing Systems*. ACM, New Orleans LA USA, 1–15. https://doi.org/10.1145/3491102.3517451
[67] Tae-Ho Lee, June-Seek Choi, and Yang Seok Cho. 2012. Context modulation of facial emotion perception differed by individual difference. *PLOS one* 7, 3 (2012), e32987.
[68] Hao Li, Laura Trutoiu, Kyle Olszewski, Lingyu Wei, Tristan Trutna, Pei-Lun Hsieh, Aaron Nicholls, and Chongyang Ma. 2015. Facial performance sensing head-mounted display. *ACM Transactions on Graphics (ToG)* 34, 4 (2015), 1–9.
[69] Jie Li, Vinoba Vinayagamoorthy, Julie Williamson, David A Shamma, and Pablo Cesar. 2021. Social VR: A new medium for remote communication and collaboration. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–6.
[70] Jinghuai Lin and Marc Erich Latoschik. 2022. Digital body, identity and privacy in social virtual reality: A systematic review. *Frontiers in Virtual Reality* 3 (2022), 974652.
[71] Chao Ma, Nianxin Guo, Faraday Davies, Yantian Hou, Suyan Guo, and Xun Zhu. 2021. Facial expressions can be categorized along the upper-lower facial axis, from a perceptual perspective. *Attention, Perception, & Psychophysics* 83 (2021), 2159–2173.
[72] Fang Ma and Xueni Pan. 2022. Visual Fidelity Effects on Expressive Self-avatar in Virtual Reality: First Impressions Matter. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, Christchurch, New Zealand, 57–65. https://doi.org/10.1109/VR51125.2022.00023
[73] Mariana Magalhães, António Coelho, Miguel Melo, and Maximino Bessa. 2024. Measuring users' emotional responses in multisensory virtual reality: A systematic literature review. *Multimedia Tools and Applications* 83, 14 (2024), 43377–43417.
[74] Meeri Mäkäräinen, Jari Kätsyri, and Tapio Takala. 2014. Exaggerating Facial Expressions: A Way to Intensify Emotion or a Way to the Uncanny Valley? *Cognitive Computation* 6, 4 (Dec. 2014), 708–721. https://doi.org/10.1007/s12559-014-9273-0
[75] Divine Maloney, Guo Freeman, and Donghee Yvette Wohn. 2020. "Talking without a Voice": Understanding Non-verbal Communication in Social Virtual Reality. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (Oct. 2020), 1–25. https://doi.org/10.1145/3415246
[76] John W. Mauchly. 1940. Significance Test for Sphericity of a Normal n-Variate Distribution. *Annals of Mathematical Statistics* 11 (1940), 204–209. https://api.semanticscholar.org/CorpusID:119821954
[77] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and inter-rater reliability in qualitative research: Norms and guidelines for CSCW and HCI practice. *Proceedings of the ACM on human-computer interaction* 3, CSCW (2019), 1–23.
[78] Rachel McDonnell and Martin Breidt. 2010. Face reality: Investigating the uncanny valley for virtual faces. In *ACM SIGGRAPH ASIA 2010 Sketches*. 1–2.
[79] Joshua McVeigh-Schultz, Anya Kolesnichenko, and Katherine Isbister. 2019. Shaping pro-social interaction in VR: an emerging design framework. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–12.
[80] Hanneke KM Meeren, Corné CRJ Van Heijnsbergen, and Beatrice De Gelder. 2005. Rapid perceptual integration of facial expression and emotional body language. *Proceedings of the National Academy of Sciences* 102, 45 (2005), 16518–16523.
[81] Elizabeth J Miller, Yong Zhi Foo, Paige Mewton, and Amy Dawel. 2023. How do people respond to computer-generated versus human faces? A systematic review and meta-analyses. *Computers in Human Behavior Reports* 10 (2023), 100283.
[82] Elizabeth J Miller, Eva G Krumhuber, and Amy Dawel. 2022. Observers perceive the Duchenne marker as signaling only intensity for sad expressions, not genuine emotion. *Emotion* 22, 5 (2022), 907.
[83] Masahiro Mori, Karl F MacDorman, and Norri Kageki. 2012. The uncanny valley [from the field]. *IEEE Robotics & automation magazine* 19, 2 (2012), 98–100.
[84] Matthew WE Murry and Derek M Isaacowitz. 2017. Age differences in emotion perception: The effects of the social environment. *International Journal of Behavioral Development* 41, 5 (2017), 597–604.
[85] Michelle B Neiss, Lindsey A Leigland, Nichole E Carlson, and Jeri S Janowsky. 2009. Age differences in perception and awareness of emotion. *Neurobiology of aging* 30, 8 (2009), 1305–1313.
[86] Lauri Nummenmaa and Manuel G Calvo. 2015. Dissociation between recognition and detection advantage for facial expressions: a meta-analysis. *Emotion* 15, 2 (2015), 243.
[87] Soo Youn Oh, Jeremy Bailenson, Nicole Krämer, and Benjamin Li. 2016. Let the Avatar Brighten Your Smile: Effects of Enhancing Facial Expressions in Virtual Environments. *PLOS ONE* 11, 9 (Sept. 2016), e0161794. https://doi.org/10.1371/journal.pone.0161794
[88] Kyle Olszewski, Joseph J Lim, Shunsuke Saito, and Hao Li. 2016. High-fidelity facial and speech animation for VR HMDs. *ACM Transactions on Graphics (TOG)* 35, 6 (2016), 1–14.
[89] Anna Orlowska, Magdalena Rychlowska, Piotr Szarota, and Eva G Krumhuber. 2023. Facial mimicry and social context affect smile interpretation. *Journal of Nonverbal Behavior* 47, 4 (2023), 471–488.
[90] Ye Pan, Shuai Tan, Shengran Cheng, Qunfen Lin, Zijiao Zeng, and Kenny Mitchell. 2024. Expressive talking avatars. *IEEE Transactions on Visualization and Computer Graphics* (2024).
[91] Ye Pan, Ruisi Zhang, Shengran Cheng, Shuai Tan, Yu Ding, Kenny Mitchell, and Xubo Yang. 2023. Emotional Voice Puppetry. *IEEE Transactions on Visualization and Computer Graphics* 29, 5 (May 2023), 2527–2535. https://doi.org/10.1109/TVCG.2023.3247101
[92] Ye Pan, Ruisi Zhang, Jingying Wang, Yu Ding, and Kenny Mitchell. 2023. Real-Time Facial Animation for 3D Stylized Character with Emotion Dynamics. In *Proceedings of the 31st ACM International Conference on Multimedia*. ACM, Ottawa ON Canada, 6851–6859. https://doi.org/10.1145/3581783.3613803
[93] Minjung Park and Hyeon-Jeong Suk. 2022. New Mobile Adaptation System for Better Avatar-mediated Communication; Facial Expressions in Memoji. In *Adjunct Publication of the 24th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, Vancouver BC Canada, 1–5. https://doi.org/10.1145/3528575.3551437
[94] Brian Parkinson. 2008. Emotions in direct and remote social interaction: Getting through the spaces between us. *Computers in Human Behavior* 24, 4 (2008), 1510–1529.
[95] Philip A Powell and Jennifer Roberts. 2017. Situational determinants of cognitive, affective, and compassionate empathy in naturalistic digital interactions. *Computers in Human Behavior* 68 (2017), 137–148.
[96] Rivu Radiah, Daniel Roth, Florian Alt, and Yomna Abdelrahman. 2023. The Influence of Avatar Personalization on Emotions in VR. *Multimodal Technologies and Interaction* 7, 4 (March 2023), 38. https://doi.org/10.3390/mti7040038
[97] René Riedl, Peter NC Mohr, Peter H Kenning, Fred D Davis, and Hauke R Heekeren. 2014. Trusting humans and avatars: A brain imaging study based on evolution theory. *Journal of Management Information Systems* 30, 4 (2014), 83–114.
[98] Ruthger Righart and Beatrice De Gelder. 2008. Rapid influence of emotional scenes on encoding of facial expressions: an ERP study. *Social cognitive and affective neuroscience* 3, 3 (2008), 270–278.
[99] Ruthger Righart and Beatrice de Gelder. 2008. Recognition of facial expressions is influenced by emotional scene gist. *Cognitive, Affective, & Behavioral Neuroscience* 8, 3 (2008), 264–272.
[100] Daniel Roth, Gary Bente, Peter Kullmann, David Mal, Chris Felix Purps, Kai Vogeley, and Marc Erich Latoschik. 2019. Technologies for social augmentations in user-embodied virtual reality. In *Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology*. 1–12.
[101] Joshua Rottman, Jasmine DeJesus, and Heather Greenebaum. 2019. Developing disgust: Theory, measurement, and application. *Handbook of emotional development* (2019), 283–309.
[102] Annie Roy-Charland, Melanie Perron, Olivia Beaudry, and Kaylee Eady. 2014. Confusion of fear and surprise: A test of the perceptual-attentional limitation hypothesis with eye movement monitoring. *Cognition and Emotion* 28, 7 (2014), 1214–1222.
[103] James A Russell and Merry Bullock. 1985. Multidimensional scaling of emotional facial expressions: similarity from preschoolers to adults. *Journal of personality and social psychology* 48, 5 (1985), 1290.
[104] Nazanin Sabri, Bella Chen, Annabelle Teoh, Steven P Dow, Kristen Vaccaro, and Mai Elsherief. 2023. Challenges of moderating social virtual reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–20.

[105] Saba Safdar, Wolfgang Friedlmeier, David Matsumoto, Seung Hee Yoo, Catherine T Kwantes, Hisako Kakai, and Eri Shigemasu. 2009. Variations of emotional display rules within and across cultures: A comparison between Canada, USA, and Japan. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement* 41, 1 (2009), 1.

[106] Alexander Schäfer, Gerd Reis, and Didier Stricker. 2022. A survey on synchronous augmented, virtual, andmixed reality remote collaboration systems. *Comput. Surveys* 55, 6 (2022), 1–27.

[107] Klaus R Scherer, Heiner Ellgring, Anja Dieckmann, Matthias Unfried, and Marcello Mortillaro. 2019. Dynamic facial expression of emotion and observer inference. *Frontiers in psychology* 10 (2019), 433041.

[108] Annett Schirmer and Ralph Adolphs. 2017. Emotion Perception from Face, Voice, and Touch: Comparisons and Convergence. *Trends in Cognitive Sciences* 21, 3 (March 2017), 216–228. https://doi.org/10.1016/j.tics.2017.01.001

[109] Kelsea Schulenberg, Guo Freeman, Lingyuan Li, and Bart Knijnenburg. 2024. Does Who You Are or Appear to Be Matter?: Understanding Identity-Based Harassment in Social VR Through the Lens of (Mis) Perceived Identity Revelation. *Proceedings of the ACM on humancomputer interaction* (2024).

[110] Valentin Schwind, Katrin Wolf, and Niels Henze. 2018. Avoiding the uncanny valley in virtual character design. *interactions* 25, 5 (2018), 45–49.

[111] Hyewon Song and Beom Kwon. 2024. Facial Animation Strategies for Improved Emotional Expression in Virtual Reality. *Electronics* 13, 13 (2024), 2601.

[112] Matthew P Spackman, Bruce L Brown, and Sean Otto. 2009. Do emotions have distinct vocal profiles? A study of idiographic patterns of expression. *Cognition and Emotion* 23, 8 (2009), 1565–1588.

[113] Elena Spadoni, Marina Carulli, Maura Mengoni, Marco Luciani, and Monica Bordegoni. 2023. Empowering virtual Humans' emotional expression in the metaverse. In *International Conference on Human-Computer Interaction*. Springer, 133–143.

[114] Yilu Sun and Andrea Stevenson Won. 2021. Despite Appearances: Comparing Emotion Recognition in Abstract and Humanoid Avatars Using Nonverbal Behavior in Social Virtual Reality. *Frontiers in Virtual Reality* 2 (Sept. 2021), 694453. https://doi.org/10.3389/frvir.2021.694453

[115] Jon Swain. 2018. A hybrid approach to thematic analysis in qualitative research: Using a practical example. *Sage research methods* (2018).

[116] Aliya Tastemirova, Johannes Schneider, Leona Chandra Kruse, Simon Heinzle, and Jan vom Brocke. 2022. Microexpressions in digital humans: perceived affect, sincerity, and trustworthiness. *Electronic Markets* 32, 3 (2022), 1603–1620.

[117] Angela Tinwell, Mark Grimshaw, Debbie Abdel Nabi, and Andrew Williams. 2011. Facial Expression of Emotion and Perception of the Uncanny Valley in Virtual Characters. *Computers in Human Behavior* 27, 2 (March 2011), 741–749. https://doi.org/10.1016/j.chb.2010.10.018

[118] Alessandro Visconti, Davide Calandra, and Fabrizio Lamberti. 2023. Comparing Technologies for Conveying Emotions through Realistic Avatars in Virtual Reality-based Metaverse Experiences. *Computer Animation and Virtual Worlds* 34, 3-4 (May 2023), e2188. https://doi.org/10.1002/cav.2188

[119] Zhibo Wang, Jingwang Ling, Chengzeng Feng, Ming Lu, and Feng Xu. 2020. Emotion-preserving blendshape update with real-time face tracking. *IEEE Transactions on Visualization and Computer Graphics* 28, 6 (2020), 2364–2375.

[120] Florian Weidner, Gerd Boettcher, Stephanie Arevalo Arboleda, Chenyao Diao, Luljeta Sinani, Christian Kunert, Christoph Gerhardt, Wolfgang Broll, and Alexander Raake. 2023. A Systematic Review on the Visualization of Avatars and Agents in AR & VR Displayed Using Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics* 29, 5 (May 2023), 2596–2606. https://doi.org/10.1109/TVCG.2023.3247072

[121] Ben James Winer, Donald R Brown, Kenneth M Michels, et al. 1971. *Statistical principles in experimental design*. Vol. 2. McGraw-Hill New York.

[122] R. Michael Winters, Bruce N. Walker, and Grace Leslie. 2021. Can You Hear My Heartbeat?: Hearing an Expressive Biosignal Elicits Empathy. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, Yokohama Japan, 1–11. https://doi.org/10.1145/3411764.3445545

[123] Takashi Yamada and Tomio Watanabe. 2014. Analysis of facial expressions during muscle strain using a hmd-based virtual arm wrestling system designed for facial electromyography measurement. In *2014 Proceedings of the SICE Annual Conference (SICE)*. IEEE, 989–994.

[124] Chi-Lan Yang, Shigeo Yoshida, Hideaki Kuzuoka, Takuji Narumi, and Naomi Yamashita. 2023. Affective Profile Pictures: Exploring the Effects of Changing Facial Expressions in Profile Pictures on Text-Based Communication. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–17. https://doi.org/10.1145/3544548.3581061

[125] Chit Yuen Yi, Matthew WE Murry, and Amy L Gentzler. 2016. Perception of emotional expressions in adults. *Journal of Individual Differences* (2016).

[126] Xin Yi, Ziyu Han, Xinge Liu, Yutong Ren, Xin Tong, Yan Kong, and Hewu Li. 2023. Catch my Eyebrow, Catch my Mind: Examining the Effect of Upper Facial Expressions on Emotional Recognition for VR Avatars. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 761–762.

[127] Shu Zhang, Xinge Liu, Xuan Yang, Yezhi Shu, Niqi Liu, Dan Zhang, and Yong-Jin Liu. 2021. The influence of key facial features on recognition of emotion in cartoon faces. *Frontiers in psychology* 12 (2021), 687974.