This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2025.3549579

VR Whispering: A Multisensory Approach for Private Conversations in Social Virtual Reality

Xueyang Wang*, Kewen Peng*, Chonghao Hao, Wendi Yu, Xin Yi, and Hewu Li

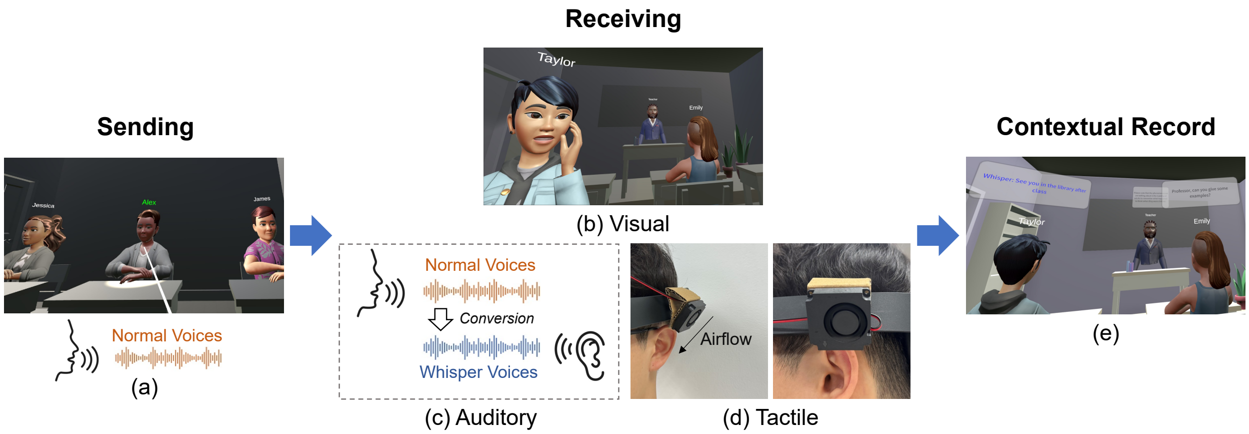


Fig. 1: Overview of the Whisper technique for private conversations in social VR. (a) Sender initiates a private voice message using a white raycast selector; ambient lighting dims and a spotlight highlights the receiver. (b-d) Multisensory interaction during message reception: (b) Visual: receiver sees sender's avatar approach and mimic whispering; (c) Audio: sender's normal voice converted to whisper; (d) Tactile: airflow near receiver's ear simulates physical presence. (e) Contextual record: displays missed public conversation content after private exchange.

Abstract— Private conversations in social Virtual Reality (VR) environments lack the nuanced cues of physical interactions, potentially diminishing the sense of privacy and social presence. This paper introduces Whisper, a novel multisensory interaction technique designed to enhance private conversations for social VR applications. We first conducted a formative study (N=20) to understand private conversation demands, limitations of existing methods, and user expectations in social VR. Informed by these insights, Whisper incorporates visual (avatar proximity, gestures and illumination), auditory (voice conversation), and tactile (simulated airflow) elements to simulate the act of whispering, providing users with an intuitive and immersive method of private communication. The technique also features a contextual record to maintain conversation continuity. We evaluated Whisper through a comparative user study (N=24) in party and classroom scenarios. Results demonstrate that Whisper significantly outperforms existing methods in sense of privacy, mode distinguishability, intimacy, perceptual realism, and social presence.

Index Terms—Whispering, private conversations, multisensory interaction, social virtual reality

1 Introduction

Social Virtual Reality (VR) has emerged as a powerful medium for digital interactions, offering immersive environments that closely mimic face-to-face communication [35]. These platforms facilitate embodied interactions between multiple users, fostering strong connections and social presence [9, 33]. As social VR applications expand into diverse domains such as digital communication, collaboration, education, and virtual workplaces [12, 31, 48, 51], the need for nuanced social interactions, including private conversations, becomes increasingly critical.

Private conversations play a vital role in both physical and virtual social contexts, enabling secure exchanges of sensitive information and minimizing disturbances to bystanders [8, 78]. These interactions, often occurring as backchannel communication alongside primary discourse [3, 55], are integral to social communication. They promote self-disclosure, relationship building, and creative innovation, while enhancing audience participation and understanding [10, 25, 50, 66].

However, current social VR applications predominantly employ a broadcasting protocol, where user speech is audible to all nearby participants. This limitation forces users to seek alternative methods for private communication, such as exploiting voice attenuation or creating separate virtual rooms. These workarounds can disrupt the user experience and potentially compromise privacy [69, 77]. While some solutions like ConeSpeech have introduced directional speech delivery [75], they often focus solely on the sender's interaction, neglecting the receiver's ability to distinguish between public and private channels. Consequently, there remains a significant need for comprehensive solutions that facilitate natural, immersive and secure private conversations in social VR environments.

To address this gap, we present Whisper, a novel multisensory technique designed to enhance private conversations in VR. Our approach is based on a formative study involving 20 experienced social VR

*These authors have contributed equally to this work.

- *Xueyang Wang, Xin Yi, and Hewu Li are with Tsinghua University and Zhongguancun Laboratory. E-mail: wang-xy22@mails.tsinghua.edu.cn, yixin@tsinghua.edu.cn, lihewu@cernet.edu.cn.*
- *Kewen Peng and Chonghao Hao are with Shanghai Jiao Tong University. E-mail: sjtu_pkw@sjtu.edu.cn, h.chchonghao@sjtu.edu.cn.*
- *Wendi Yu is with University of Sydney. E-mail: weyu0667@uni.sydney.edu.au.*
- *Xin Yi is the corresponding author.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

users, which explored user demands, evaluated existing methods, and identified key expectations for private communication in social VR. Building on these insights, Whisper incorporates visual, auditory, and tactile elements to create an immersive and realistic private conversation experience akin to real-world whispering (Figure 1).

At its core, Whisper employs visual cues using nonverbal signals [1, 42], such as proxemic metaphors and body postures, to simulate the act of whispering. This visual aspect is seamlessly coupled with an auditory transformation that converts the sender's normal voice into a whisper-like sound, enhanced by spatial audio to simulate proximity. To further heighten the realism, Whisper incorporates tactile feedback through a small, silent fan that is mounted on Head-Mounted Display (HMD) and generates subtle airflow near the receiver's ear, replicating the sensation of someone speaking closely. Complementing these sensory elements, a contextual record feature displays frontchannel conversation content after private conversations, ensuring users remain connected to the broader social context while engaging in private communication.

To validate our technique, we conducted a comprehensive user study with 24 participants, evaluating Whisper against two existing methods (text messaging and private talk channels) in two realistic scenarios: a casual party and a classroom setting. Our results demonstrated Whisper's superiority in terms of mode distinguishability, sense of privacy, intimacy, perceptual realism, and social presence. Furthermore, 17 out of 24 participants selected Whisper as their preferred method, reporting that it fostered a socially rich atmosphere and strengthened interpersonal connections.

This paper makes three key contributions: 1) A comprehensive analysis of private conversation demands, limitations of existing methods, and user expectations in social VR environments; 2) The design and implementation of Whisper, to our knowledge, the first technique to simultaneously integrate visual, auditory, and tactile elements for facilitating private conversations in social VR; 3) Empirical evidence supporting the effectiveness of our approach in enhancing the user experience of private conversations in social VR.

2 Related Work

2.1 Social Interaction in Virtual Reality

Social Virtual Reality (VR) enables remote users to interact in shared virtual environments using immersive technologies like head-mounted displays [35]. Platforms such as VRChat, Horizon Workrooms, Rec Room, and Roblox offer diverse activities including conversations, world creation, gaming, and collaborative work [31, 65]. Studies show that social VR supports meaningful interactions, enhances social skills, and yields positive psychological effects [4, 42, 80].

A key feature of social VR is its embodied interaction through customizable avatars. Advanced tracking technologies map users' movements, expressions, and gaze onto these virtual representations [45, 68], enabling social experiences that closely mirror face-to-face interactions [1, 79]. This embodied control proves particularly effective for non-verbal communication, a crucial component of social interaction.

The quality of these virtual interactions significantly influences social presence—the "sense of being with another" [5]. Both the interactivity of interpersonal activities and the realism of behavioral representations contribute to presence levels [26, 70]. These findings emphasize the importance of designing interactions that enhance social presence, particularly for intimate communications like private conversations in VR environments.

2.2 Private Communication in Digital Spaces

Private communication, or backchannel communication, involves interactions visible only to specific participants [8]. In digital spaces, it serves various purposes unsuitable for public discourse, such as personal discussions during meetings or private assistance in group activities [50, 55]. While digital private communication includes methods like instant messaging and email [3], these are not ideal for immersive social VR environments.

Current social VR platforms typically use broadcast methods for voice communication, lacking dedicated private features. ConeSpeech [75] proposed directional speech delivery but doesn't address how receivers distinguish between public and private speech. Studies on private communication in video conferencing [14, 27] and hybrid meetings [47] don't directly apply to embodied social VR. Our work aims to fill this gap by designing a comprehensive private communication experience for social VR.

Researchers have developed silent speech technologies to address privacy concerns. SilentVoice [17] captures soft speech through ingressive airflow, while PrivateTalk [76] uses hand-to-mouth gestures. Amazon introduced a novel interaction feature for Alexa: a whisper mode [52]. In this mode, when a user whispers to Alexa, the device responds by whispering back at a reduced volume. However, these approaches may be challenging in social VR due to reduced audibility and unnatural speaking methods. Our approach allows users to speak normally, offering comfort and clarity while maintaining privacy in social VR settings.

2.3 Proxemics and Attention Guidance in Multi-User VR

Interpersonal distance dynamics are crucial in non-verbal communication [72]. Hall's proxemics theory defines four zones of social intimacy [21], which has been applied to social VR design [73]. Research has explored behavioral differences between co-located and remote VR participants [53], social virtual augmentations [56], and collaboration patterns [74]. However, these studies haven't addressed proxemic cues in private VR communication. Our research innovatively introduce proximity-based interactions for private communications in social VR.

Guiding user attention in VR environments is another challenge [40]. Techniques range from arrows [38] to halos [20], with diegetic methods gaining popularity [49]. Lee et al. [34] combined light and spatial audio cues to direct attention in group conversations. However, existing methods don't systematically address attention guidance and mode distinction in private conversations. Our approach uses lighting and spatial audio synergistically to create a private, immersive experience without exposing users' behaviors.

2.4 Multisensory Interfaces in VR

Multisensory VR interfaces enhance users' sense of presence [59]. While visual and auditory stimuli are common [29], other sensory modalities remain underdeveloped [46]. Recent research has explored tactile sensations in VR, particularly using airflow and ultrasound for facial and oral stimulation [22, 23].

Practical approaches integrate tactile hardware with HMDs. Head Mounted Wind [7] uses fans to blow air onto the face, VaIR [54] employs a pneumatic system, and FaceHaptics [71] includes various tactile modules. Shen et al. developed ultrasonic transducers for mouth haptics [63]. However, these works overlook the ear as a haptic target, despite its sensitivity to tactile stimulation [2].

Our work addresses this gap by introducing ear-focused tactile stimulation for private VR conversations. We use fan-generated airflow to simulate whispering breath, enhancing privacy and intimacy in virtual social interactions. Importantly, our thin components could be integrated into future headsets in a practical and consumer-friendly way [63]. Our approach contributes to developing more nuanced, multisensory communication systems in VR.

3 Formative Study: Exploring Design Considerations for Private Conversations in VR

To inform the design of an enhanced private conversation method for social VR, we conducted a semi-structured interview study with 20 experienced VR users. This study aimed to understand the issues of current methods and key factors enhancing privacy and user experience in virtual private conversations, addressing the unique challenges and opportunities presented by social VR platforms.

3.1 Participants

We recruited a diverse group of participants to ensure a comprehensive perspective on private conversations in VR. The study included 11 active social VR users and 9 graduate students from design schools with experience in technology and design. This combination provided a balance between practical user experience and design insights. Participants ranged in age from 18 to 28 years (M = 23.3, SD = 3.28), with an equal gender distribution (10 males, 10 females). Twelve participants had over a month of VR experience, with six reporting more than a year. Eleven had prior exposure to social VR platforms like VRChat or participated in VR social meetings and interaction games. Among these, six had attempted private conversations in VR environments. This varied experience level ensured a rich dataset encompassing both novice and expert perspectives. The study was approved by Science and Technology Ethics Committee of Tsinghua University, and all participants provided signed informed consent.

3.2 Study Setup and Procedure

3.2.1 Videos for Eliciting User Experiences in Social VR

Adopting a video-stimulated approach [34], we used two videos to stimulate discussion and gather feedback on current social VR practices and private conversation methods.

Video 1: Social VR Scenarios showcased common VR social scenarios, including lectures, conferences, multiplayer collaborations, chats, and informal social settings with collaborative elements (e.g., movie theaters, concerts, travel, KTV, and multiplayer games). This video aimed to immerse participants in a shared social VR experience, providing a common reference point for discussion.

Video 2: Current Private Conversation in VR illustrated existing private communication methods in social VR, based on online videos and existing surveys [37]. It featured *Text Message* (sending private messages), *Private Rooms* (setting up personal chatrooms), *Proximity Talking* (adjusting message reception by moving closer or changing seats), *Private Talk* (establishing private communication channels), and *External Devices* (using devices like cellphones outside the VR for private talks), demonstrating the transition between public and private modes in various scenarios.

Both videos incorporated content from popular social VR platforms such as VRChat, Horizon Workrooms, Rec Room, BigScreen, and Roblox, as referenced in [37,65]. This selection ensured comprehensive coverage of common social VR applications, providing participants with a consistent and representative social VR experience to discuss.

3.2.2 Procedure

The interview process, illustrated in Figure 2, consisted of sequential video viewings followed by question-and-answer sessions. After each 5-10 minute video, participants answered questions about their experiences, feelings, and needs regarding socializing and private conversations in VR. The first set of questions focused on demands for private conversations under various social VR scenarios, while the second set evaluated existing private conversation methods and explored participants' expectations for ideal solutions.

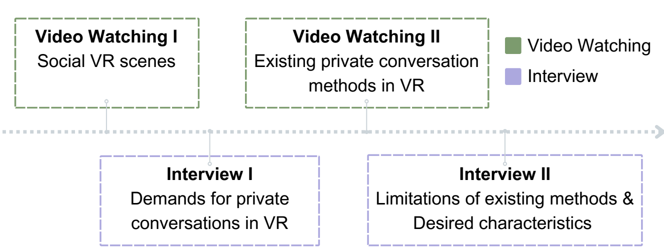


Fig. 2: Formative study procedure: Participants watched two videos sequentially, each followed by an interview session to gather insights on private conversation demands and expectations.

Interviews were conducted via VooVMeeting, simulating remote communication similar to VR while offering greater accessibility than VR headsets. Each study session lasted approximately 45 minutes, providing ample time for in-depth discussion and reflection.

3.3 Results

3.3.1 Demands for Private Conversation in VR

Participants expressed a clear need for private conversations across various VR contexts, mirroring real-life social interactions. The most frequently mentioned scenarios included VR lectures (60%), multi-user collaborations (60%), conferences (55%), and general chatting (55%). Importantly, the requirements for private conversations varied depending on the specific VR scenario. For instance, in VR lectures or conferences, the priority was often to exchange messages discreetly without being noticed (P2, P3, P10, P15). In contrast, during collaborations or casual chatting, the focus shifted towards not disturbing others in the virtual space (P3, P5, P11, P14, P15, P17, P18). This variability highlights the need for flexible private communication solutions that can adapt to different social contexts within VR environments.

3.3.2 Limitations of Existing Techniques

Despite the long-standing development of social VR, current private communication techniques exhibit several limitations:

1. *Misalignment with VR Interaction Paradigms:* Some methods that work well in real-life settings are less suited for VR socializing. For example, many users reported frustrating experiences with typing in VR for *Text Messages* (P7, P16, P18), noting that voice communication feels more natural and efficient in virtual environments. Similarly, maintaining a low voice in *Proximity Talking* was described as "a bit tiring" (P15), indicating a mismatch with VR's immersive nature.

2. *Privacy Issue:* A key challenge in private conversations is the risk of exposure. *Proximity Talking*, while providing a sense of presence, introduces risks of unintended participants overhearing due to unclear directionality or control over conversation boundaries. Additionally, physical movements in VR to adjust proximity can draw unwanted attention, as noted by eight participants.

3. *Disruption of VR Immersion:* Managing the balance between private conversations and the surrounding public environment is challenging. *Private Rooms* allow for secluded discussions but can disrupt the flow of interaction by detaching users from their original group (P3), potentially causing them to miss important information (P5). *External Devices* completely disconnect users from the VR experience, significantly breaking immersion.

4. *Mode Switching Difficulties:* The transition between private and public modes in *Text Message*, *Private Rooms*, and *Private Talk* can be prone to operational errors (P5, P6, P7), affecting the smoothness of mode transitions, risking in privacy leakage, and ultimately disrupting the overall VR experience.

These limitations underscore the need for a more integrated, VR-native approach to private conversations that maintains immersion, ensures privacy, and aligns with natural interaction paradigms in virtual environments.

3.3.3 Desired Characteristics

Based on participant feedback, we identified five key characteristics expected in an ideal private conversation solution for social VR:

1. Convenient Interaction: Users emphasized the need for simple, intuitive methods to initiate and engage in private conversations. Complicated operations can deter users from utilizing private conversation features (P11). The ability to easily and accurately choose conversation partners and exchange information is crucial. Additionally, maintaining a smooth and inclusive flow in private conversations is essential, requiring both the avoidance of distractions and the flexibility to respond to them when necessary.

2. Privacy Protection: Participants highlighted two main aspects of privacy in VR conversations: content security and conversation discretion. Users want assurance that their conversation content cannot be accessed or overheard by unintended parties, and that the very act of engaging in a private conversation remains unnoticed by others, both in virtual and physical worlds (P4, P5, P15). This dual-layer privacy concern reflects the unique challenges of maintaining confidentiality in shared virtual spaces.

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2025.3549579

3. Mode Transparency: Clear indicators of conversation mode (public vs. private) are crucial to reduce cognitive load and prevent accidental information leakage. Participants expressed a need for unambiguous signals or recognition mechanisms to indicate when a private conversation is active. Additionally, they wanted clear identification of their current conversation partner in private mode, minimizing the need for constant system checks and enhancing overall user experience.

4. Seamless Immersion: Maintaining VR immersion during private conversations is vital for both speakers and listeners. From the speaker's perspective, natural, voice-based communication is preferred over typing, aligning with the immersive nature of VR. Listeners benefit from multi-modal elements in VR private conversations, including visual feedback, auditory voice effects, and tactile sensations simulating close proximity (P2, P9, P16, P18). This comprehensive sensory input enhances the receiver's ability to fully perceive and interpret the sender's message, mitigating the "information bias," (P2, P6, P9, P17, P18, P20), thereby enhancing the overall communication experience.

5. Situational Awareness: Users emphasized the importance of maintaining awareness of their public environment while engaged in private conversations. This dual awareness enables users to stay connected with the larger group dynamics and respond appropriately to public prompts, preventing awkward social situations (P3, P12). The ability to seamlessly balance private communications with ongoing public interactions is crucial for a socially adept and immersive VR experience.

4 Whisper: Multisensory Private Conversations for Social VR

Our formative study revealed diverse user requirements for private conversations in VR, presenting significant design challenges. Inspired by the observation that 40% of participants mentioned using whispers in real-life gatherings, we developed Whisper, a multisensory technique for private communication in social VR. Whispering, a form of unvoiced speech, serves as an intuitive and natural mode of private communication among humans [36], making it an ideal design metaphor for VR environments.

While physical whispers are limited to co-located individuals, virtual environments offer greater flexibility in manipulating space and avatar behavior [73]. Leveraging this advantage, Whisper addresses key characteristics identified in our study: ease of interaction, privacy and confidentiality, mode transparency, seamless immersion, and situational awareness. These features align with familiar social behaviors while providing an inherent sense of privacy in the virtual space.

Figure 1 provides an overview of the Whisper interaction, demonstrating how it translates real-world whispering behavior into a VR context. Figure 3 illustrates its technical components.

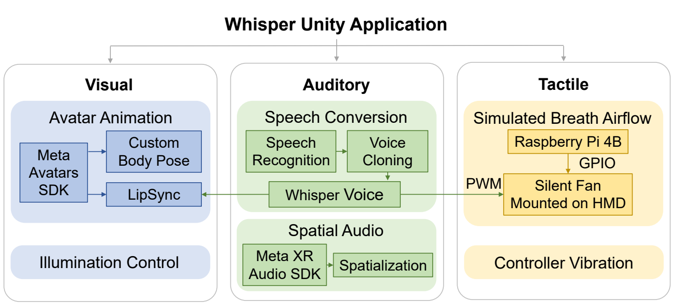


Fig. 3: Technical components of the Whisper Unity application, showcasing the integration of visual, auditory, and tactile elements for enhanced private conversations in social VR.

4.1 Visual Components

Whisper's visual elements are designed to enhance both the sender's and receiver's experience, clearly delineating private conversations and fostering a sense of intimacy in social VR.

4.1.1 Sender's Interface

While public communication in VR typically involves direct voice broadcasting to nearby users, private conversations require a more deliberate interaction approach. For initiating private conversations, we adopted a controller raycast-based directional selection method, inspired by ConeSpeech [75], and incorporated light variations to provide an intuitive and non-intrusive signal for the initiation and conclusion of private conversations (Figure 1(a)). The interaction process unfolds as follows: 1) The user selects the intended recipient using a white raycast controlled by the VR controller; 2) Upon pressing the controller trigger, ambient lighting dims, and a spotlight illuminates the chosen recipient; 3) The user speaks in their normal voice, which is transmitted as a private message; 4) Releasing the trigger ends the message and restores normal lighting. This method combines precise recipient selection with subtle environmental feedback to clearly distinguish private communications from public discourse.

4.1.2 Receiver's Experience

The receiver's visual experience comprises two main features: avatar animation and illumination control.

(a) First-person View

(b) Top-down view

Fig. 4: Visual representation of a Whisper message from the receiver's perspective. (a) First-person view: The sender's avatar (Taylor) approaches the receiver, moving to whispering proximity with a hand raised to simulate covering the mouth. (b) Top-down view: Spatial rearrangement during a private conversation, showing the sender's avatar movement and gradual dimming of ambient lighting with spotlight emphasis.

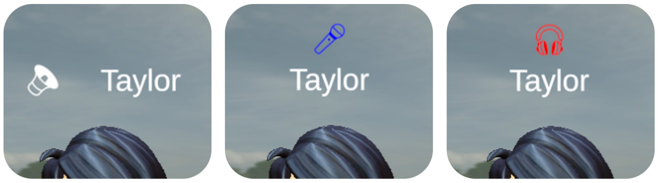
Avatar Animation Inspired by real-world whispering behavior, we implemented proximity-based interactions (Figure 4). The sender's avatar moves to a distance of 0.5 meters from the receiver, based on Hall's proxemics model [21], balancing intimacy and comfort. To accommodate the limited field of view in HMDs, the sender's avatar appears at a 45-degree angle in front of the receiver, ensuring easy identification without excessive head movement.

The avatar's movement lasts one second, utilizing an ease-in-out function for visual comfort:

$$t' = t^2(3 - 2t) \tag{1}$$

where t is the normalized time (0 to 1), and t' is the eased value used for interpolation.

To create an authentic whispering experience while maintaining privacy, we carefully designed the receiver's visual feedback using the Meta Avatars SDK. When receiving a private message, the receiver observes custom-designed animations that mimic natural whispering behavior, including subtle body poses, lip-sync movements, and a hand-covering-mouth gesture. These visual cues, along with the status indicators shown in Figure 5, help receivers clearly distinguish between private and public communications.



(a) Public speech (b) Sending whisper (c) Receiving whisper

Fig. 5: Visual indicators for different speaking state.

A key privacy-preserving feature of our design is that all feedback is exclusively visible to the receiver. Bystanders observe no indication that a private conversation is taking place, following the inherent characteristics of physical whispering [36] and established social acceptability principles [32]. This selective visibility ensures private communications remain discreet while maintaining an engaging experience for the participating users.

Illumination Control Drawing from research on creating intimate virtual spaces [41], we incorporated dynamic lighting adjustments. As shown in Figure 4(b), ambient lighting gradually dims to 50% during the sender's avatar movement, with a spotlight emphasizing the sender. This lighting change further distinguishes private conversations from public interactions and enhances the sense of confidentiality and intimacy.

4.2 Auditory Components

Audio design is crucial in virtual environments, second only to vision in its importance for interaction and perception [29]. To emulate the unique phonetic qualities of whispering, we developed two key audio features:

4.2.1 Voice Conversion

Considering the potential negative effects of prolonged whisper voice use on vocal cords [17, 57], we implemented a speech conversion technique allowing users to speak normally while producing whisper-like output. This process involves:

1) Initial voice sample collection: Users record a 20-30 second whisper speech sample for voice cloning algorithm training.

2) Real-time conversion: Using ByteDance's Volcano Engine[1], we employ speech recognition coupled with voice cloning to transform the user's normal voice into a whisper in real-time.

While this approach enhances privacy perception and maintains speaker characteristics, it introduces some latency. Our tests with 50 speech samples (3-5 seconds each) revealed average latencies of 0.91 seconds (SD=0.27) for streaming speech recognition and 2.09 seconds (SD=0.54) for voice cloning. As speech processing technologies advance, we anticipate this issue will be mitigated. To mitigate potential confusion caused by these delays, we implemented visual indicators to represent different speaking states (Figure 5).

4.2.2 Spatial Audio

To enhance the immersive experience, we leveraged the Meta XR Audio SDK to implement spatial audio features. This technology allows users to localize audio sources in three-dimensional space, creating a more realistic auditory environment [29]. When receiving a private message, the sender's audio source is virtually positioned near the receiver's ear, simulating the close proximity of physical whispering. We also employed near-field rendering to approximate acoustic diffraction effects for close audio sources, further enhancing realism. To ensure clarity of private messages, we reduce ambient sound volume to 50% during whisper reception.

[1]https://www.volcengine.com

4.3 Tactile Components

Whisper features a simulated airflow of breath, mimicking the breathy quality of physical whispering [36]. To simulate this sensation, we mounted a silent fan ($3cm \times 3cm \times 1cm$) on the HMD (Figure 1(d)), controlled by a Raspberry Pi 4B. This fan generates subtle airflow near the user's ear, providing a crucial tactile dimension to the virtual whispering experience. The fan's close noise is about 20 decibels, equivalent to calm breathing, ensuring it doesn't interfere with the audio experience.

The fan speed is dynamically adjusted using Pulse Width Modulation (PWM) based on the real-time volume of the whispered message to simulate speaking exhalation:

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N} s_i^2} \tag{2}$$

$$DutyRatio = \min(RMS \times ScalingRatio, 100) \tag{3}$$

where s_i are the audio samples, N is the number of samples (512 in our implementation), and ScalingRatio is a constant (empirically set to 5000) to produce an airflow intensity closely resembling human breath during speech.

Additionally, the user's controller vibrates for one second when receiving a private message, simulating potential physical contact during a real-world whisper.

4.4 Contextual Record Module

To address users' desire for contextual awareness during private conversations, we developed the Contextual Record feature (Figure 1(e)). This feature records and transcribes public conversations occurring in the virtual environment during private message exchanges. Upon completion of a private conversation, text bubbles appear above the respective speakers' avatars, displaying transcribed public conversations in standard text and private messages in blue text. These bubbles remain visible for 10 seconds, providing users a brief window to catch up on missed public conversations.

5 Evaluation Study

Building upon the insights from our formative study (Section 3) and the development of Whisper (Section 4), we conducted a systematic evaluation to assess Whisper's effectiveness in addressing the identified gaps in VR private conversation methods.

5.1 Study Design

We employed a mixed-factorial design, with *Technique* as a within-subjects factor and *Scenario* as a between-subjects factor. This approach allowed us to compare Whisper against two baseline techniques inspired by commercial social VR platforms: *Text Message* and *Private Talk* (Figure 7).

- *Private Talk*, modeled after VIVRESE[2]—an immersive collaboration platform for HTC VIVE, creates a private communication channel with a specific person, using "Hold" and "Resume" keys to switch between private and public modes.
- *Text Message*, common in platforms like VRChat and Rec Room, allows users to send private messages. To align with user preferences from our formative study, we implemented speech recognition as the input method for text messages, enhancing comparability with voice-based techniques.

For the *Scenario* factor, we selected two representative settings frequently mentioned in our formative study: *Party* (social VR gatherings) and *Class* (lecture environment). These scenarios represent distinct social dynamics—*Party* as a multi-focus, equality-matched environment, and *Class* as a single-focus, authority-ranked setting [15]. Each scenario featured 6-7 avatars, including 4 NPCs, 2 participants, and an additional professor in the *Class* setting. Avatar positioning reflected

[2]https://business.vive.com

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2025.3549579

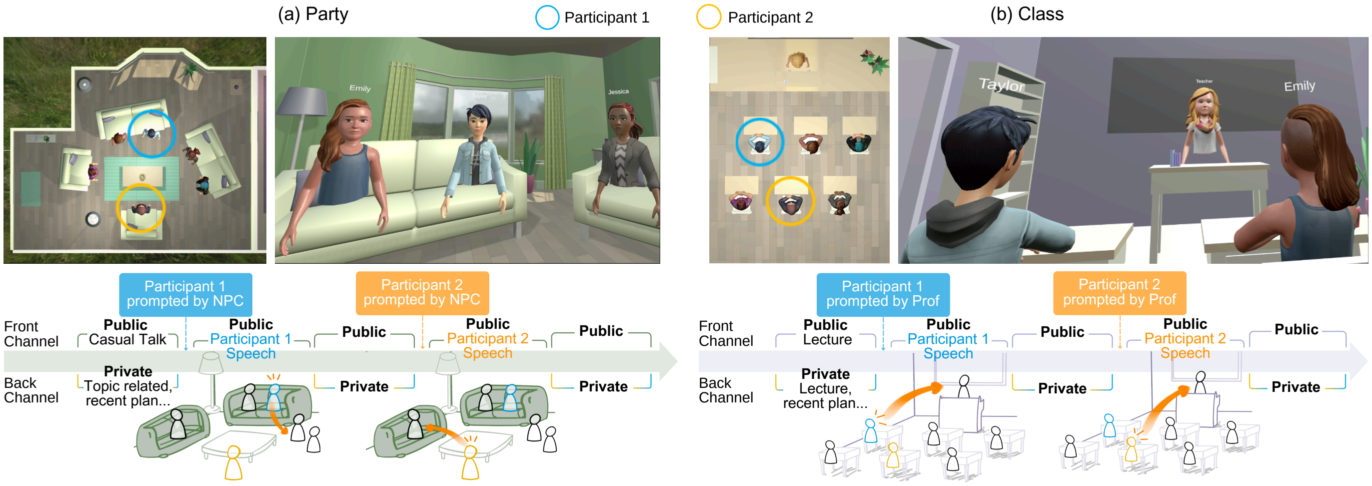


Fig. 6: Experimental scenarios and procedure for evaluation study: (a) Party setting with casual conversations; (b) Class setting with a lecture format. Both scenarios alternate between public and private channels, with NPCs prompting participant interactions.

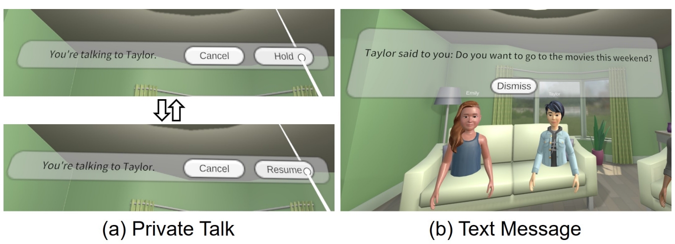


Fig. 7: Baseline techniques for private communication in VR: (a) Private Talk, featuring a dedicated channel with "Hold" and "Resume" controls for mode switching; (b) Text Message, showing a speech-to-text interface for sending private messages.

typical arrangements for each scenario: face-to-face in the *Party* and different rows in the *Class* (Figure 6).

To ensure a fair comparison across techniques, we created a dynamic social environment that balanced controlled interactions with realistic scenarios [34]. NPCs engaged in turn-based public discourse using pre-recorded audio—lectures for *Class* and casual conversations for *Party*. Additionally, to mimic natural social dynamics, each NPC randomly initiated one-on-one interactions with participants, posing topic-related questions.

Participants were tasked with navigating this multi-layered communication environment. They were encouraged to engage in frequent private conversations using the technique being tested, discussing either scenario-related topics or personal matters. Simultaneously, they were required to maintain awareness of the public channel, responding promptly and accurately when prompted by NPCs. This setup, as depicted in Figure 6, allowed us to effectively evaluate users' experiences in switching between public and private modes of communication.

5.2 Measurements

Our evaluation focused on two primary areas: user preferences for private conversations in VR and the immersive experience of social VR techniques. We derived eight key metrics to address both user expectations identified in our formative study (e.g., privacy, distinguishability, convenience, situational awareness) and crucial aspects of effective private conversations in VR environments (e.g., clarity, intimacy, realism). These metrics included: 1) *Sense of Privacy*: feeling of confidentiality during private conversations; (2) *Mode Distinguishability*: clarity between public and private conversation modes; 3) *Convenience*: ease of use for initiating and maintaining private conversations; 4) *Information Clarity*: clear and accurate transmission of messages; 5) *Situation Awareness*: ability to maintain awareness of surroundings while in private mode; 6) *Intimacy*: feelings of closeness during private conversations; 7) *Perceptual Realism*: resemblance to real-life private communication experiences; 8) *Overall Satisfaction*: general user satisfaction with the private communication method.

To assess users' sense of immersion, we used adapted versions of the Igroup Presence Questionnaire (IPQ) [61] for self-presence and the Networked Minds Social Presence Inventory (NMSPI) [24] for social presence. Additionally, we employed the User Experience Questionnaire (UEQ) [60] to evaluate Whisper specifically, gaining deeper insights into its user experience. All questionnaires used a 7-point Likert scale to capture nuanced participant feedback.

5.3 Participants and Apparatus

We recruited 12 pairs of participants (24 individuals total), such as close friends or couples, based on the understanding that private conversations often occur within close relationships [36]. Participants included 14 females and 10 males, aged 21 to 33 years (M = 24.63, SD = 3.49), all with normal or corrected vision. The majority (87.5%) had prior VR experience, with an average interest in private conversations in social VR rated at 5.33 out of 7 (SD = 0.85). The study was approved by Science and Technology Ethics Committee of Tsinghua University, and all participants provided signed informed consent before participating in the study.

We implemented the virtual environments using Unity 2022.3.36f1, employing cartoon-style avatars from the Meta Avatar SDK to align with mainstream social VR platforms. This design choice enhances the generalizability of our findings to current social VR applications. We utilized Photon Fusion 2[3] for real-time multiplayer networking and Photon Voice 2 for high-quality, low-latency voice chat. The application ran on ASUS laptops and was streamed to Meta Quest Pro headsets.

5.4 Procedure

The study began with a comprehensive overview and a two-minute instructional video on the three techniques. Participants recorded a 25-second whisper-like audio sample for Whisper's model training before entering VR. Seated in separate, quiet rooms, they were fitted with HMDs and fans, ensuring comfort and proper setup.

Participants then experienced each of the three techniques sequentially in their assigned scenario (*Party* or *Class*), with the order randomized using Latin square sampling. They alternated between being listeners and speakers, engaging in private conversations while also interacting with the scenario-specific prompts from NPCs. This design

[3]https://www.photonengine.com

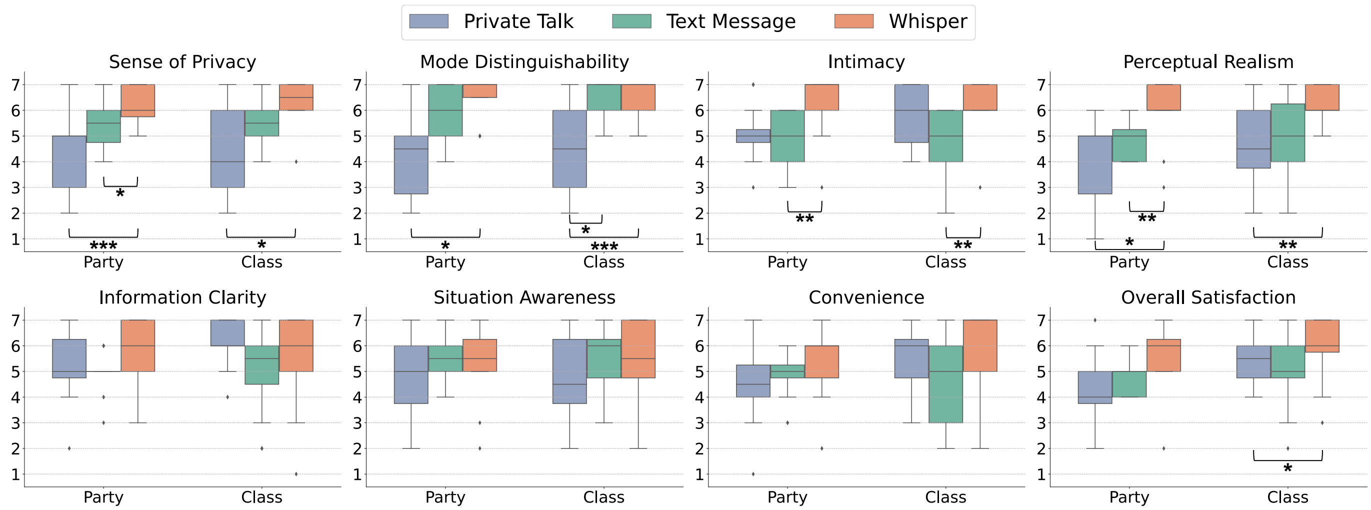


Fig. 8: Users' subjective ratings across eight metrics for private conversation techniques in VR. Box plots show median, quartiles, and outliers for Party and Class scenarios. Significance levels: *$p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

allowed participants to fully experience each technique in a realistic social VR context.

After completing the experiment, participants were asked to fill out assessment scales and engage in brief interviews to provide insights into the reasoning behind their assessment outcomes. The entire experiment lasted approximately 60 minutes.

5.5 Analysis

All 24 participants actively engaged in the study, providing a rich dataset for analysis. Given the participants' consistent engagement and accurate responses to NPC prompts, we included all 24 questionnaire responses in our quantitative analysis. To evaluate the techniques, we employed a mixed ANOVA with the Aligned Rank Transform (ART) [13] for non-parametric factorial analysis. This method accounted for the non-normal distribution typical of Likert data. We treated *Technique* as a within-subject factor and *Scenario* as a between-subject factor, applying Mauchly's test for sphericity and Greenhouse-Geisser corrections where necessary. Post-hoc comparisons used Bonferroni-corrected paired t-tests to identify specific differences while controlling for multiple comparisons. We consistently used $p < 0.05$ as the significance threshold. In reporting results (Section 5.6), we indicated the specific significance level ($p < 0.05$, $p < 0.01$, $p < 0.001$) to provide more detailed information about the strength of our findings.

For qualitative data from interviews, we conducted thematic analysis [44], coding responses into key themes to complement our quantitative findings and provide deeper insights into user experiences and preferences.

5.6 Results

5.6.1 Quantitative Analysis of Users' Subjective Ratings

Our analysis revealed significant main effects of *Technique* across multiple metrics, with no significant effects of *Scenario* or *Technique-Scenario* interactions, indicating the robustness of our findings across different social contexts.

Whisper demonstrated superior performance across several key metrics for private conversations (Figure 8). In terms of *Sense of Privacy* ($F(2,44) = 16.00, p < 0.001$), Whisper outperformed Private Talk in both scenarios (Party: $p < 0.001$, Class: $p < 0.05$) and Text Message in the Party scenario ($p < 0.05$). *Mode Distinguishability* ($F(2,44) = 18.15, p < 0.001$) favored Whisper over Private Talk in both scenarios (Party: $p < 0.05$, Class: $p < 0.001$) and over Text Message in the Class scenario ($p < 0.05$). Whisper excelled in *Intimacy* ($F(2,44) = 12.00, p < 0.001$), surpassing Text Message in both scenarios (both $p < 0.01$). For *Perceptual Realism* ($F(2,44) = 14.44, p < 0.001$), Whisper outperformed Private Talk in both scenarios (Party: $p < 0.05$, Class: $p < 0.01$) and Text Message in the Party scenario ($p < 0.01$). While not statistically significant, Whisper achieved comparable or higher levels of Information Clarity, Situation Awareness, and Convenience. *Overall Satisfaction* ($F(2,44) = 5.18, p < 0.05$) showed Whisper's superiority over Private Talk in the Class scenario ($p < 0.05$).

Presence measurements further underscored Whisper's effectiveness (Figure 9). For self-presence, Whisper demonstrated significant improvements across all metrics. In *Spatial Presence* ($F(2,44) = 14.50, p < 0.001$), Whisper outperformed Text Message in both scenarios (both $p < 0.01$). For *Involvement* ($F(1.6,34.8) = 18.70, p < 0.001$), Whisper surpassed both alternatives in the Party scenario (Private Talk: $p < 0.01$, Text Message: $p < 0.001$) and Text Message in the Class scenario ($p < 0.01$). *Contextual Realism* ($F(2,44) = 22.59, p < 0.001$) showed Whisper outperforming Text Message in both scenarios (Party: $p < 0.05$, Class: $p < 0.01$) and Private Talk in the Class scenario ($p < 0.001$). These results suggest that Whisper creates a more immersive and realistic private conversation experience, enhancing users' sense of being present in the virtual environment.

Social presence measures revealed similar trends, highlighting Whisper's ability to foster interpersonal connections. In *Co-Presence* ($F(2,44) = 9.54, p < 0.001$), Whisper excelled over Text Message in both scenarios (both $p < 0.01$). *Perceived Attentional Engagement* ($F(2,44) = 8.71, p < 0.01$) showed Whisper outperforming both alternatives in the Party scenario (Private Talk: $p < 0.05$, Text Message: $p < 0.01$). For *Perceived Emotional Contagion* ($F(2,44) = 13.80, p < 0.001$), Whisper surpassed Text Message in both scenarios (both $p < 0.01$). Whisper also showed advantages in *Perceived Comprehension* and *Perceived Behavioral Interdependence*, although the differences were less pronounced. These findings indicate that Whisper enhances users' ability to perceive and engage with others during private conversations, potentially leading to more meaningful and emotionally resonant interactions.

The User Experience Questionnaire (UEQ) results (Figure 10) further complemented these findings, with participants rating Whisper highly across various dimensions. Notably, Whisper excelled in enjoyability, interest, pleasantness, innovation, and leading-edge design. High ratings in ease of use and efficiency suggest that Whisper successfully balances novelty with usability, a critical factor for the adoption of new VR technologies.

5.6.2 Qualitative Insights from User Interviews

Thematic analysis of user interviews revealed strong preference for Whisper, with 17 out of 24 participants selecting it as their preferred private communication method. This clear preference reflects Whis-

This article has been accepted for publication in IEEE Transactions on Visualization and Computer Graphics. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TVCG.2025.3549579

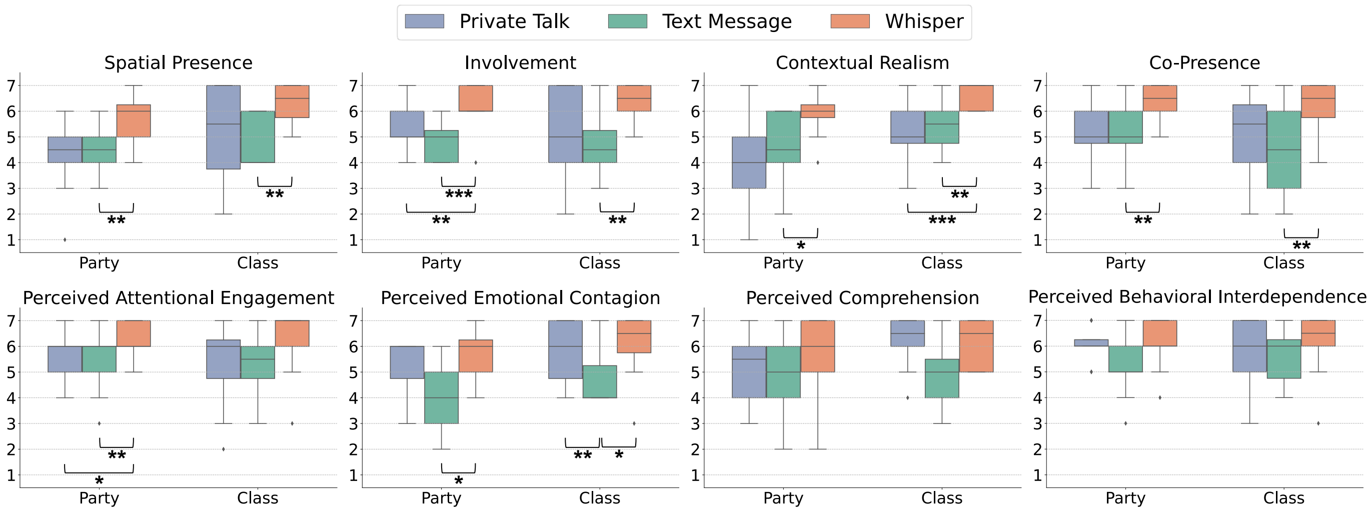


Fig. 9: Self-presence and social presence metrics across techniques and scenarios. Box plots show median, quartiles, and outliers. Significance levels: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

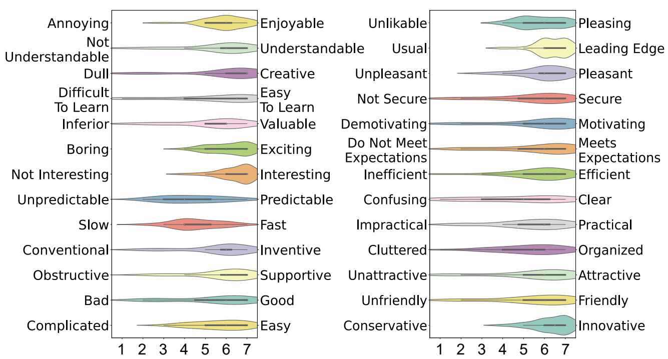


Fig. 10: UEQ results for Whisper. Violin plots show response distributions across bipolar scales, with wider sections indicating higher response frequency. Horizontal lines represent median values.

per's successful balance of privacy, realism, and usability in social VR environments. Our analysis identified several key themes that illuminate users' experiences and preferences:

Privacy and Mode Distinction: Whisper emerged as the preferred method for ensuring privacy, with 15 out of 24 participants rating it highest. Users praised its multisensory design for creating a strong sense of confidentiality. P8 noted, *"The speaker's body language and focused lighting gave me a strong psychological cue of privacy."* The discreet interaction method of Whisper was also commended, as P12 explained: *"I only need to operate the controller to enter a private conversation, without spatial movement. Bystanders would hardly notice, thinking I've just fallen silent."* Sixteen participants found Whisper's feedback facilitated easy mode differentiation, enhancing confidence during private conversations. In contrast, Private Talk was criticized for its ambiguity in privacy assurance and high risk of mode confusion.

Usability and Clarity: Twelve users praised Whisper for its operational simplicity, criticizing text input and establishing private talk channels as comparatively complex. Despite some latency in voice conversion, only two participants reported impacts on communication efficiency. Notably, six users found Whisper provided better clarity than Text Message, preserving crucial voice elements like intonation and emotion.

Immersion and Social Presence: Whisper significantly enhanced the sense of immersion and social presence. Eighteen users favored its rich multisensory feedback, with 11 participants noting its similarity to real-life whispering. P13 commented that Whisper *"recreated that feeling of sneakily talking to classmates."* Whisper's close-proximity interactions and tactile feedback were particularly effective in creating a socially rich atmosphere. P7 remarked, *"Seeing the other person turn towards me... made me feel instantly closer,"* while P13 vividly described the tactile experience: *"The breath airflow near my ear is fantastic... gave me goosebumps."* In contrast, Text Message and Private Talk were likened to texting and phone calls, perceived as less immersive and detached from the face-to-face VR scenario.

Negative Feedback and Failure Cases: User interviews revealed several limitations in the current implementation, primarily concerning voice conversion and initial user adaptation. P21 and P23 reported that converted whispers were occasionally too quiet for clear comprehension, necessitating reliance on the Contextual Record feature and reducing communication efficiency. P10 noted issues with voice distortion affecting speaker recognition, while P5 suggested implementing adjustable whisper volume as a potential solution. First-time users also experienced temporary discomfort with the technique's novel elements: P7 and P9 were startled by sudden avatar movements, while P1, P5, P14, and P18 reported initial unease with the tactile feedback. These adaptation challenges led to suggestions for incorporating an orientation phase to help users acclimate to these novel interactive features before engaging in actual conversations.

6 Discussion

In this paper, based on user's expectation from a formative study (Section 3), developed and evaluated Whisper, a multisensory private communication technique for social VR. In this section, we further interpret our results and discuss key considerations for the practical implementation of private communication in social VR environments.

6.1 Contributions of Multisensory Design

Advantages of Multisensory Interface in Social VR: When designing interactive experiences for social VR, consideration of the environment's immersive nature is crucial, as users engage in activities that mirror face-to-face interactions [35]. Existing solutions like ConeSpeech, PrivateTalk, and text messaging typically transplant conventional digital communication features into VR platforms without fully leveraging VR's unique affordances [9,33]. This approach often fails to recreate authentic interpersonal interactions [68], detaching users from the immersive face-to-face context. In contrast, Whisper's multisensory interface authentically recreates physical whispering characteristics through synchronized feedback channels, producing significantly higher levels of presence, realism, and immersion compared to baseline methods, as validated in Section 5.6.

Complementary Role of Sensory Modalities: Our evaluation revealed how different sensory channels work together to create an effective private communication experience. Visual cues emerged as fundamental privacy indicators through multiple elements. Gestures (P11, P20) effectively conveyed emotional expression and acknowledgment [67], while avatar proximity (P6, P8, P10, P20) created a sense of intimacy and privacy [39]. Strategic illumination variations (P1, P12, P13) contributed to establishing an intimate atmosphere, aligning with research on the psychological effects of lighting [19]. These visual elements collectively enhanced privacy perception and social presence in VR [43, 64]. The auditory component served dual purposes: enhancing presence and clearly signaling communication modes (P13, P16, P17, P20). Some research also suggests that synchronous auditory feedback not only increases presence and realism [30] but also improves environmental memory [11]. Complementing these elements, tactile feedback through simulated airflow created convincing private conversation experiences (P9, P12, P20, P22, P23) by adding physical authenticity [18] and enriching the multimodal experience [11]. This integration of multiple sensory channels demonstrates how each modality contributes unique yet complementary aspects to create a cohesive and immersive private communication experience.

6.2 Scalability for Devices and Contexts

While Whisper demonstrates promising potential for private VR communication, its current implementation presents scalability challenges, primarily due to hardware dependencies. The reliance on silent fans for tactile feedback may limit widespread adoption across VR platforms. However, this hardware investment can be justified by its multi-purpose potential—the same tactile infrastructure could enhance other VR experiences, such as weather simulation [28]. Future iterations could explore more adaptable solutions, including software-based alternatives and different tactile mechanisms like vibration or ultrasonic feedback [63], potentially offering more cost-effective implementations while maintaining the essential tactile experience.

The technique's appropriateness also varies across social contexts and relationships. Drawing parallels with physical whispering behaviors [8, 36], Whisper may be more suitable for interactions between familiar individuals and casual, entertainment-oriented settings (P2, P6, P13). We also recommend several other promising application scenarios, such as virtual theaters and libraries where quiet communication is essential. For formal contexts like business meetings, we recommend implementing selected visual features (e.g., lighting modulation or gestures) to maintain professionalism while preserving privacy benefits. This contextual adaptability ensures Whisper can serve diverse communication needs while respecting established social norms.

6.3 Social Acceptability

Our design carefully considers social acceptability principles [32], balancing interaction effectiveness with social appropriateness. Whisper's interface builds on familiar VR controller interactions, minimizing the learning curve while ensuring discreet operation (P23). The minimal button operation and absence of visible signals to bystanders reduce the likelihood of drawing unwanted attention or causing misinterpretation (P11). Additionally, the contextual records feature helps participants maintain awareness of public conversations while engaged in private exchanges (P5, P10, P15, P24), further supporting smooth social integration (P6, P17).

Further enhancements could extend Whisper's social acceptability in dynamic VR environments. For instance, implementing an inbox mechanism for managing multiple private messages would better support scenarios where parallel conversations naturally occur, such as social gatherings or collaborative workspaces (P15, P16, P21). This approach would maintain Whisper's discreet nature while improving its flexibility in handling complex social situations.

6.4 Ethical Considerations

Understanding and mitigating emerging harassment in social VR remains a critical topic [6, 16]. We advocate for commercial social VR platforms to implement robust permission mechanisms when adopting private conversation methods like Whisper to reduce harassment risks [16]. For instance, speakers could be required to send a request for private conversation, which the listener must accept before communication can begin. Users should have the option to enable a "Do Not Disturb" mode when they wish to avoid private messages or interruptions. Additionally, users should have the right to blacklist harassers and report them to the platform. While preserving user privacy, platforms could request permission to identify whisper content to filter out malicious messages, including personal attacks, hate speech, or discriminatory comments.

For Whisper's voice conversion feature, platforms should obtain user consent for voice recording and processing, paying particular attention to data security to prevent identity-based attack risks [62]. Platforms could even offer voice conversion as a user option within voice modulators, using machine-generated voices to maintain user anonymity [58]. These measures aim to create a safer, more controlled environment for private communications in VR, balancing the benefits of enhanced interaction with necessary safeguards against misuse.

7 Limitations and Future Work

Although Whisper demonstrates promising potential for enhancing private conversations in social VR, our study revealed several limitations that warrant future investigation. From a technical perspective, the current implementation lacks certain physical characteristics of natural whispering, such as temperature and humidity components in the tactile feedback. The technique's usability could also be enhanced through alternative interaction methods beyond controller-based input, such as proximity triggers or gesture recognition, while maintaining privacy. Additionally, while we expect continued improvements in voice processing technology to reduce conversion latency, additional design features are needed to better manage these delays.

Our experimental design, while allowing for controlled comparisons, had limitations in ecological validity. The use of script-driven NPCs, one-on-one interactions, and fixed user positions may not fully capture the complexity of real-world social VR scenarios. Future studies should test Whisper in more dynamic, multi-user environments across various contexts, such as educational settings and professional meetings. This approach would provide a more comprehensive evaluation of Whisper's performance in authentic social situations. Additionally, expanding the age distribution of participants would mitigate potential bias towards new technologies among younger users.

Future validation should also incorporate third-party observers to assess the discretion of private conversations from an uninvolved perspective. This evaluation is crucial for understanding Whisper's real-world effectiveness and its impact on broader social dynamics in VR spaces. Long-term usage studies are also necessary to reveal insights into user adaptation, potential misuse scenarios, and the technology's influence on social norms in virtual environments.

8 Conclusion

This paper introduces Whisper, a novel multisensory technique designed to facilitate private conversations in social Virtual Reality environments. Inspired by real-world whispering behavior, Whisper integrates visual, auditory, and tactile elements to create an immersive and intuitive private conversation experience. Our user studies demonstrated that Whisper significantly outperforms existing methods in terms of sense of privacy, mode distinguishability, intimacy, and perceptual realism. Moreover, Whisper enhanced both self-presence and social presence, contributing to more engaging and naturalistic interactions in virtual spaces. By closely mimicking physical world interactions, Whisper bridges the gap between digital and face-to-face communication, potentially opening new avenues for social interaction in virtual environments.

Acknowledgments

This work was supported by the Natural Science Foundation of China under Grant No. 62132010, 62472243, and 2022YFB3105201.

References

[1] N. Aburumman, M. Gillies, J. A. Ward, and A. F. d. C. Hamilton. Nonverbal communication in virtual reality: Nodding as a social signal in virtual interactions. *International Journal of Human-Computer Studies*, 164:102819, 2022. 2

[2] R. Ackerley, I. Carlsson, H. Wester, H. Olausson, and H. Backlund Wasling. Touch perceptions across skin sites: differences between sensitivity, direction discrimination and pleasantness. *Frontiers in behavioral neuroscience*, 8:54, 2014. 2

[3] C. Atkinson. *The backchannel: how audiences are using Twitter and social media and changing presentations forever*. New Riders, 2009. 1, 2

[4] M. Barreda-Ángeles and T. Hartmann. Psychological benefits of using social virtual reality platforms during the covid-19 pandemic: The role of social and spatial presence. *Computers in Human Behavior*, 127:107047, 2022. 2

[5] F. Biocca, C. Harms, and J. K. Burgoon. Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence: Teleoperators & virtual environments*, 12(5):456–480, 2003. 2

[6] L. Blackwell, N. Ellison, N. Elliott-Deflo, and R. Schwartz. Harassment in social virtual reality: Challenges for platform governance. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–25, 2019. 9

[7] S. Cardin, D. Thalmann, and F. Vexo. Head mounted wind. In *proceeding of the 20th annual conference on Computer Animation and Social Agents (CASA2007)*, pp. 101–108, 2007. 2

[8] S. Cogdill, T. L. Fanderclai, J. Kilborn, and M. G. Williams. Backchannel: whispering in digital conversation. In *Proceedings of the 34th Annual Hawaii International Conference on System Sciences*, pp. 8–pp. IEEE, 2001. 1, 2, 9

[9] J. J. Cummings and A. Shore. Distinguishing social virtual reality: Comparing communication channels across perceived social affordances, privacy, and trust. *Computers in Human Behavior*, p. 108427, 2024. 1, 8

[10] A. R. Dennis, J. A. Rennecker, and S. Hansen. Invisible whispering: Restructuring collaborative decision making with instant messaging. *Decision Sciences*, 41(4):845–886, 2010. 1

[11] H. Dinh, N. Walker, L. Hodges, C. Song, and A. Kobayashi. Evaluating the importance of multi-sensory input on memory and the sense of presence in virtual environments. In *Proceedings IEEE Virtual Reality (Cat. No. 99CB36316)*, pp. 222–228, 1999. doi: 10.1109/VR.1999.756955 9

[12] E. Dzardanova, V. Kasapakis, D. Gavalas, and S. Sylaiou. Virtual reality as a communication medium: a comparative study of forced compliance in virtual reality versus physical world. *Virtual Reality*, 26(2):737–757, 2022. 1

[13] L. A. Elkin, M. Kay, J. J. Higgins, and J. O. Wobbrock. An aligned rank transform procedure for multifactor contrast tests. In *The 34th annual ACM symposium on user interface software and technology*, pp. 754–768, 2021. 7

[14] T. Erickson, N. S. Shami, W. A. Kellogg, and D. W. Levine. Synchronous interaction among hundreds: An evaluation of a conference in an avatar-based virtual environment. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 503–512, 2011. 2

[15] A. P. Fiske. *Structures of social life: The four elementary forms of human relations: Communal sharing, authority ranking, equality matching, market pricing.* Free Press, 1991. 5

[16] G. Freeman, S. Zamanifard, D. Maloney, and D. Acena. Disturbing the peace: Experiencing and mitigating emerging harassment in social virtual reality. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW1):1–30, 2022. 9

[17] M. Fukumoto. Silentvoice: Unnoticeable voice input by ingressive speech. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, pp. 237–246, 2018. 2, 5

[18] J. K. Gibbs, M. Gillies, and X. Pan. A comparison of the effects of haptic and visual feedback on presence in virtual reality. *International Journal of Human-Computer Studies*, 157:102717, 2022. 9

[19] R. Gifford. Light, decor, arousal, comfort and communication. *Journal of environmental psychology*, 8(3):177–189, 1988. 9

[20] U. Gruenefeld, A. E. Ali, S. Boll, and W. Heuten. Beyond halo and wedge: Visualizing out-of-view objects on head-mounted virtual and augmented reality devices. In *Proceedings of the 20th international conference on human-computer interaction with mobile devices and services*, pp. 1–11, 2018. 2

[21] E. T. Hall. The hidden dimension. *Garden City*, 1966. 2, 4

[22] P.-H. Han, Y.-S. Chen, C.-E. Hsieh, H.-C. Wang, and Y.-P. Hung. Hapmosphere: Simulating the weathers for walking around in immersive environment with haptics feedback. In *2019 IEEE World Haptics Conference (WHC)*, pp. 247–252. IEEE, 2019. 2

[23] P.-H. Han, Y.-S. Chen, K.-C. Lee, H.-C. Wang, C.-E. Hsieh, J.-C. Hsiao, C.-H. Chou, and Y.-P. Hung. Haptic around: multiple tactile sensations for immersive environment and interaction in virtual reality. In *Proceedings of the 24th ACM symposium on virtual reality software and technology*, pp. 1–10, 2018. 2

[24] C. Harms and F. Biocca. Internal consistency and reliability of the networked minds measure of social presence. In *Seventh annual international workshop: Presence*, vol. 2004. Universidad Politecnica de Valencia Valencia, 2004. 6

[25] D. Harry, J. Green, and J. Donath. Backchan. nl: integrating backchannels in physical space. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 1361–1370, 2009. 1

[26] F. Herrera, S. Y. Oh, and J. N. Bailenson. Effect of behavioral realism on social interactions inside collaborative virtual environments. *Presence*, 27(2):163–182, 2020. 2

[27] E. Hu, M. A. R. Azim, and S. Heo. Fluidmeet: Enabling frictionless transitions between in-group, between-group, and private conversations during virtual breakout meetings. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 1–17, 2022. 2

[28] F. Hülsmann, J. Fröhlich, N. Mattar, and I. Wachsmuth. Wind and warmth in virtual reality: implementation and evaluation. In *Proceedings of the 2014 virtual reality international conference*, pp. 1–8, 2014. 9

[29] F. Immohr, G. Rendle, C. Kehling, A. Lammert, S. Göring, B. Froehlich, and A. Raake. Subjective evaluation of the impact of spatial audio on triadic communication in virtual reality. In *2024 16th International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 262–265. IEEE, 2024. 2, 5

[30] A. C. Kern and W. Ellermeier. Audio in vr: Effects of a soundscape and movement-triggered step sounds on presence. *Frontiers in Robotics and AI*, 7:20, 2020. 9

[31] S. Kimmel, F. Jung, A. Matviienko, W. Heuten, and S. Boll. 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*, pp. 1–16, 2023. 1, 2

[32] M. Koelle, S. Ananthanarayan, and S. Boll. Social acceptability in hci: A survey of methods, measures, and design strategies. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, 19 pages, p. 1–19. Association for Computing Machinery, New York, NY, USA, 2020. 5, 9

[33] E. Kukshinov, D. Harley, K. Szita, R. Hadi Mogavi, C. Macarthur, and L. E. Nacke. Disembodied, asocial, and unreal: How users reinterpret designed affordances of social vr. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*, pp. 1914–1925, 2024. 1, 8

[34] G. Lee, D. Y. Lee, G.-M. Su, and D. Manocha. "may i speak?": Multi-modal attention guidance in social vr group conversations. *IEEE Transactions on Visualization and Computer Graphics*, 30(5):2287–2297, May 2024. doi: 10.1109/tvcg.2024.3372119 2, 3, 6

[35] J. Li, V. Vinayagamoorthy, J. Williamson, D. A. Shamma, and P. Cesar. Social vr: A new medium for remote communication and collaboration. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–6, 2021. 1, 2, 8

[36] X. Li. Whispering: the murmur of power in a lo-fi world. *Media, Culture & Society*, 33(1):19–34, 2011. 4, 5, 6, 9

[37] J. Limbago et al. Designing user-centric private conversation methods in the metaverse. Master's thesis, 2023. 3

[38] Y.-C. Lin, Y.-J. Chang, H.-N. Hu, H.-T. Cheng, C.-W. Huang, and M. Sun. Tell me where to look: Investigating ways for assisting focus in 360 video. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 2535–2545, 2017. 2

[39] A. M. Lomanowska and M. J. Guitton. Spatial proximity to others determines how humans inhabit virtual worlds. *Computers in Human Behavior*, 28(2):318–323, 2012. doi: 10.1016/j.chb.2011.09.015 9

[40] A. MacQuarrie and A. Steed. Cinematic virtual reality: Evaluating the effect of display type on the viewing experience for panoramic video. In *2017 IEEE Virtual Reality (VR)*, pp. 45–54. IEEE, 2017. 2

[41] D. Maloney and G. Freeman. Falling asleep together: What makes activities in social virtual reality meaningful to users. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*, pp. 510–521, 2020. 5

[42] D. Maloney, G. Freeman, and D. Y. Wohn. " talking without a voice" understanding non-verbal communication in social virtual reality. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2):1–25, 2020. 2

[43] D. Maloney, G. Freeman, and D. Y. Wohn. "talking without a voice": Understanding non-verbal communication in social virtual reality. 4(CSCW2), article no. 175, 25 pages, Oct. 2020. doi: 10.1145/3415246 9

[44] N. McDonald, S. Schoenebeck, and A. Forte. 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):1–23, 2019. 7

[45] J. McVeigh-Schultz, A. Kolesnichenko, and K. Isbister. Shaping pro-social interaction in vr: an emerging design framework. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pp. 1–12, 2019. 2

[46] M. Melo, G. Gonçalves, P. Monteiro, H. Coelho, J. Vasconcelos-Raposo, and M. Bessa. Do multisensory stimuli benefit the virtual reality experience? a systematic review. *IEEE transactions on visualization and computer graphics*, 28(2):1428–1442, 2020. 2

[47] Q. Mu, M. Borowski, J. E. S. Grønbæk, S. Bødker, and E. Hoggan. Whispering through walls: towards inclusive backchannel communication in hybrid meetings. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–16, 2024. 2

[48] S. Mystakidis, E. Berki, and J.-P. Valtanen. Deep and meaningful e-learning with social virtual reality environments in higher education: A systematic literature review. *Applied Sciences*, 11(5):2412, 2021. 1

[49] L. T. Nielsen, M. B. Møller, S. D. Hartmeyer, T. C. Ljung, N. C. Nilsson, R. Nordahl, and S. Serafin. Missing the point: an exploration of how to guide users' attention during cinematic virtual reality. In *Proceedings of the 22nd ACM conference on virtual reality software and technology*, pp. 229–232, 2016. 2

[50] S. Nobarany, M. Haraty, S. S. Fels, and B. D. Fisher. Leveraging trust relationships in digital backchannel communications. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, pp. 1579–1584, 2011. 1, 2

[51] H. Park, D. Ahn, and J. Lee. Towards a metaverse workspace: opportunities, challenges, and design implications. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–20, 2023. 1

[52] E. Parviainen and M. L. J. Søndergaard. Experiential qualities of whispering with voice assistants. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pp. 1–13, 2020. 2

[53] I. Podkosova and H. Kaufmann. Co-presence and proxemics in shared walkable virtual environments with mixed colocation. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–11, 2018. 2

[54] M. Rietzler, K. Plaumann, T. Kränzle, M. Erath, A. Stahl, and E. Rukzio. Vair: Simulating 3d airflows in virtual reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 5669–5677, 2017. 2

[55] C. Ross, M. Terras, C. Warwick, and A. Welsh. Enabled backchannel: Conference twitter use by digital humanists. *Journal of documentation*, 67(2):214–237, 2011. 1, 2

[56] D. Roth, C. Klelnbeck, T. Feigl, C. Mutschler, and M. E. Latoschik. Beyond replication: Augmenting social behaviors in multi-user virtual realities. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 215–222. IEEE, 2018. 2

[57] A. D. Rubin, V. Praneetvatakul, S. Gherson, C. A. Moyer, and R. T. Sataloff. Laryngeal hyperfunction during whispering: reality or myth? *Journal of voice*, 20(1):121–127, 2006. 5

[58] N. Sabri, B. Chen, A. Teoh, S. P. Dow, K. Vaccaro, and M. Elsherief. Challenges of moderating social virtual reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–20, 2023. 9

[59] M. V. Sanchez-Vives and M. Slater. From presence to consciousness through virtual reality. *Nature reviews neuroscience*, 6(4):332–339, 2005. 2

[60] M. Schrepp, J. Thomaschewski, and A. Hinderks. Construction of a benchmark for the user experience questionnaire (ueq). 2017. 6

[61] T. Schubert, F. Friedmann, and H. Regenbrecht. The experience of presence: Factor analytic insights. *Presence: Teleoperators & Virtual Environments*, 10(3):266–281, 2001. 6

[62] K. Schulenberg, G. Freeman, L. Li, and B. Knijnenburg. 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. 9

[63] V. Shen, C. Shultz, and C. Harrison. Mouth haptics in vr using a headset ultrasound phased array. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2022. 2, 9

[64] H. J. Smith and M. Neff. Communication behavior in embodied virtual reality. CHI '18, 12 pages, p. 1–12. Association for Computing Machinery, New York, NY, USA, 2018. doi: 10.1145/3173574.3173863 9

[65] P. Sykownik, L. Graf, C. Zils, and M. Masuch. The most social platform ever? a survey about activities motives of social vr users. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 546–554, 2021. doi: 10.1109/VR50410.2021.00079 2, 3

[66] P. Sykownik, D. Maloney, G. Freeman, and M. Masuch. Something personal from the metaverse: goals, topics, and contextual factors of self-disclosure in commercial social vr. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 1–17, 2022. 1

[67] T. J. Tanenbaum, N. Hartoonian, and J. Bryan. "how do i make this thing smile?": An inventory of expressive nonverbal communication in commercial social virtual reality platforms. CHI '20, 13 pages, p. 1–13. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3313831.3376606 9

[68] J. Tham, A. H. Duin, L. Gee, N. Ernst, B. Abdelqader, and M. McGrath. Understanding virtual reality: Presence, embodiment, and professional practice. *IEEE Transactions on Professional Communication*, 61(2):178–195, 2018. 2, 8

[69] M. Vondráček, I. Baggili, P. Casey, and M. Mekni. Rise of the metaverse's immersive virtual reality malware and the man-in-the-room attack & defenses. *Computers & Security*, 127:102923, 2023. 1

[70] C. Wienrich, K. Schindler, N. Döllinqer, S. Kock, and O. Traupe. Social presence and cooperation in large-scale multi-user virtual reality-the relevance of social interdependence for location-based environments. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 207–214. IEEE, 2018. 2

[71] A. Wilberz, D. Leschtschow, C. Trepkowski, J. Maiero, E. Kruijff, and B. Riecke. Facehaptics: Robot arm based versatile facial haptics for immersive environments. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2020. 2

[72] J. Williamson, J. Li, V. Vinayagamoorthy, D. A. Shamma, and P. Cesar. Proxemics and social interactions in an instrumented virtual reality workshop. In *Proceedings of the 2021 CHI conference on human factors in computing systems*, pp. 1–13, 2021. 2

[73] J. R. Williamson, J. O'Hagan, J. A. Guerra-Gomez, J. H. Williamson, P. Cesar, and D. A. Shamma. Digital proxemics: Designing social and collaborative interaction in virtual environments. In *Proceedings of the 2022 CHI conference on human factors in computing systems*, pp. 1–12, 2022. 2, 4

[74] E. Wong, J. Sánchez Esquivel, G. Leiva, J. E. S. Grønbæk, and E. Velloso. Practice-informed patterns for organising large groups in distributed mixed reality collaboration. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–18, 2024. 2

[75] Y. Yan, H. Liu, Y. Shi, J. Wang, R. Guo, Z. Li, X. Xu, C. Yu, Y. Wang, and Y. Shi. Conespeech: Exploring directional speech interaction for multi-person remote communication in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2647–2657, 2023. 1, 2, 4

[76] Y. Yan, C. Yu, Y. Shi, and M. Xie. Privatetalk: Activating voice input with hand-on-mouth gesture detected by bluetooth earphones. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*, pp. 1013–1020, 2019. 2

[77] S. Yang and W. Zhang. Presence and flow in the context of virtual reality storytelling: What influences enjoyment in virtual environments? *Cyberpsychology, Behavior, and Social Networking*, 25(2):101–109, 2022. 1

[78] N. Yankelovich, J. McGinn, M. Wessler, J. Kaplan, J. Provino, and H. Fox. Private communications in public meetings. In *CHI'05 extended abstracts on Human factors in computing systems*, pp. 1873–1876, 2005. 1

[79] A. Yassien, P. ElAgroudy, E. Makled, and S. Abdennadher. A design space for social presence in vr. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society*, pp. 1–12, 2020. 2

[80] S. Zamanifard and G. Freeman. "the togetherness that we crave" experiencing social vr in long distance relationships. In *Companion Publication of the 2019 Conference on Computer Supported Cooperative Work and Social Computing*, pp. 438–442, 2019. 2