

OpenMic: Utilizing Proxemic Metaphors for Conversational Floor Transitions in Multiparty Video Meetings

Erzhen Hu
University of Virginia
Charlottesville, VA, USA
eh2qs@virginia.edu

Austin Houck
University of Virginia
Charlottesville, VA, USA
akh4rf@virginia.edu

Jens Emil Grønbæk
Aarhus University
Aarhus, Denmark
jensemil@cs.au.dk

Seongkook Heo
University of Virginia
Charlottesville, VA, USA
seongkook@virginia.edu

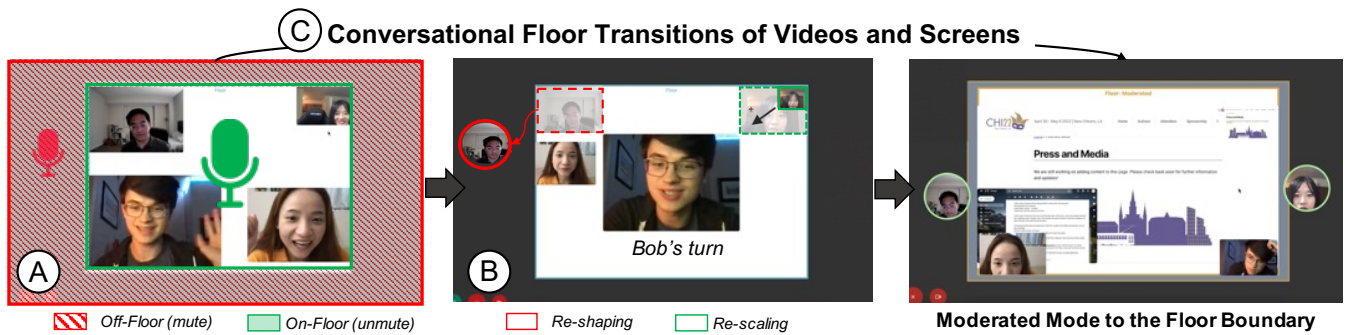


Figure 1: OpenMic provides two main features that incorporate proxemic metaphors: A) First, *Virtual Floor* acts as a virtual conversational floor that mediates turn-taking and microphone management. Second, *Malleable Mirrors* support users in B) changing the position and size of their video to control perceived distance to others and C) continuously reconfiguring videos and screens to support conversational floor transitions on and around the floor.

ABSTRACT

Turn-taking is one of the biggest interactivity challenges in multiparty remote meetings. One contributing factor is that current videoconferencing tools lack support for proxemic cues; i.e., spatial cues that humans use to enact their social relations and intentions. While more recent tools provide support for proxemic metaphors, they often focus on approach and leave-taking rather than turn-taking. In this paper, we present OpenMic, a videoconferencing system that utilizes proxemic metaphors for conversational floor management by providing 1) a Virtual Floor that serves as a fixed-feature space for users to be aware of others' intention to talk, and 2) Malleable Mirrors, which are video and screen feeds that can be continuously moved and resized for conversational floor transitions. Our exploratory user study found that these system features can aid the conversational flow in multiparty video meetings. With this work, we show potential for embedding proxemic metaphors to

support conversational floor management in videoconferencing systems.

CCS CONCEPTS

• Human-centered computing → Collaborative and social computing systems and tools.

KEYWORDS

multiparty video meeting, video conferencing, turn-taking, non-verbal cues, proxemics

ACM Reference Format:

Erzhen Hu, Jens Emil Grønbæk, Austin Houck, and Seongkook Heo. 2023. OpenMic: Utilizing Proxemic Metaphors for Conversational Floor Transitions in Multiparty Video Meetings. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*, April 23–28, 2023, Hamburg, Germany. ACM, New York, NY, USA, 17 pages. <https://doi.org/10.1145/3544548.3581013>

1 INTRODUCTION

Turn-taking is a core challenge in the support for remote meetings that has been a topic of research for decades [66, 75]. In a recent study on remote work during the COVID-19 pandemic, turn-taking was reported to be the main interactivity challenge in video-mediated communication [61], with participants expressing

difficulties dealing with overlapping and interruptive talk. While there are many contributing factors to these turn-taking issues, one problem lies in impoverished and reduced non-verbal cues, such as gestures and head/body movements, in video meetings, as well as the lack of a shared frame of reference for interpreting these cues [13, 14, 53, 57, 75].

To address this issue, prior research has explored socio-spatial (proxemic) perspectives for considering how video-conferencing technology and office spaces together shape people's opportunities for enacting their social relations [13, 37, 43, 57, 65, 67]. Moreover, research on Media Spaces and Social VR systems shows that specialized setups can support turn-taking based on people's natural head turns and mutual gaze [57, 65] or through virtual 3D avatar embodiment [10, 12, 25, 78, 80]. However, such systems require specialized equipment, which is often neither available nor desirable in the current trend of remote work that relies on everyday devices such as laptops [15, 54, 60]. Another line of research and commercial platforms focuses on utilizing 2D or 3D proximity spaces with user representation of game-like avatars [1, 3] or live video windows [5–7, 9] and/or spatial audio. Gonzalez Dias et al. [24] coin the notion of Conversational Transitions (CTs) to articulate how systems support users in managing their engagement in conversations through approaching and leave-taking as well as pre-, post- and during meetings. However, it remains to be investigated how proxemic metaphors can be used for conversational transitions at the level of *turn-taking dynamics* in meetings (i.e., *within* the conversation rather than *in and out of* conversations).

In this work, we propose to support conversational floor management through proxemic metaphors. In particular, we propose to support continuous transitions with people's video windows (Fig. 1) rather than binary transitions (e.g., a button for virtually raising the hand, or muting/unmuting) in video-conferencing tools. The conversational floor is a cognitively shared attentive space for turn-taking [58]. We propose to materialize the conversational floor as a user interface element around which turn-taking motifs can be enacted through the visual transitions with video windows. To explore such interaction principles for supporting turn-taking, we developed OpenMic – a video-conferencing system with a 2D virtual space wherein users can spatially organize their video and screen-share windows in relation to a Virtual Floor. To implement this concept for virtual meetings, OpenMic has two main features, the Virtual Floor and Malleable Mirrors (Figure 1). The Virtual Floor provides a *fixed-feature space* [35] for conveying the intention to talk and defining a boundary for managing conversations. Malleable Mirrors [31] provide proximity-based interactions (akin to, e.g., [1, 5, 7, 9]) but with the additional feature of adjustable video size. Relative image size of people is known to serve as a visual cue for perceived proximity [18, 27] in video-mediated communication. This combination of features together enables what we call *Conversational Floor Transitions*, i.e., manipulations of video and screen feeds to support continuous rather than binary transitions for taking or ceding the floor in remote group conversations.

We conducted an exploratory study with two deployments to analyze patterns in participants' *Conversational Floor Transitions* and better understand their role in turn-taking with OpenMic. We found that users exploit different types of interpersonal proximity when addressing groups vs. individuals, and that the 2D space

on and around the Virtual Floor was partitioned into zones that serve different functions for smooth turn transitions by mediating gradual engagement with others. From our results, we draw three main implications for the design of video-conferencing interfaces to support conversational floor management.

- (1) *Support a variety in group sizes.* Our results show that the current Virtual Floor design was most useful for 8-person meetings and less for 4-person meetings. Yet, making the floor more configurable may allow for adapting to different group sizes.
- (2) *Find a balance between free and curated positioning of videos.* The ability to manipulate video windows introduces trade-offs. While it supports non-verbal means for conveying the intention to talk, it may also add additional effort to the interaction.
- (3) *Support different kinds of proximity cues.* Analysing video window interaction patterns, we found that relative position was used to address individuals whereas relative size was used to address groups.

With this work, we make the following contributions: 1) the novel concept of Conversational Floor Transitions based on proxemic metaphors for turn-taking in multiparty video meetings; 2) a prototype of a video-conferencing system (OpenMic) that enables conversational floor transitions; 3) a user study in which we identify turn-taking patterns, turn-taking zones, and design trade-offs in free vs. curated positioning of video windows; 4) implications for the design of video-conferencing features for conversational floor management.

2 BACKGROUND AND RELATED WORK

We provide an overview of prior research on turn-taking in video communication, proxemics, and related user interface solutions for multiparty remote meetings.

2.1 Proxemics and Collaborative Interfaces

Proxemics is a social theory originating from E.T. Hall [35] concerned with how spatial relations (e.g., distance and orientation) between people and objects in the environment enable and hinder people's opportunities for enacting their social relations. Ideas of proxemics are often applied to HCI and particularly research on Media Spaces and video-conferencing tools [13, 14, 24, 37, 43, 50, 57, 66, 67]. We draw on the following three notions from the proxemics theory to design new mechanisms for turn-taking in multiparty video meetings.

2.1.1 Fixed/Semifixed-feature Space. E.T. Hall's notions of *fixed* and *semifixed* features have inspired several HCI works on co-located collaboration [28–30, 43, 51]. The notions articulate how features such as walls (fixed) and furniture (semifixed) in the environment serve to frame our opportunities for social interaction [35]. Prior work has used them as entities for triggering interface responses through gradual engagement as a function of proximity [28, 48, 49], incorporated tables and walls in cross-device sharing techniques [29], and articulated how furniture is used for proxemic

transitions [30]. Video-mediated communication research has emphasized the socially configuring role of office furniture in video-conference room design [43, 57]. This line of work has inspired the idea of the *Virtual Floor* in OpenMic, which “furnishes” the virtual 2D space with a fixed feature to which participants can orient their attention.

2.1.2 Perceptual Cues for Proximity. E.T. Hall’s work is best known for the concept of proxemic distance zones, i.e., people move in close proximity to others to show intention to engage with them, and increase the distance to others when signaling to disengage [35]. These spatial patterns have been shown to re-occur consistently in VR spaces with 3D avatars [78], and even in virtual 2D spaces. For example, the *relative position between 2D avatars* provides means for serendipitous encounters by “bumping” into each other [44].

Interestingly, other perceptual cues have shown to impact people’s perception of proxemic distance in video-mediated communication [18, 27]. Grayson and Coventry [27] found that in video-mediated communication, image size was a key factor in determining impressions of distance in photographs: the larger an image is, the closer it appears, regardless of the object-to-background ratio. Furthermore, during instructional videoconferencing, Ellis [18] found that perceived proximity to the teacher affected the student’s performance in the course and attitude towards the teacher.

2.2 Remote Communication Interfaces with Proxemic Cues

We categorize remote communication systems into three different approaches to providing proxemic cues: media spaces, virtual 3D spaces, and virtual 2D spaces.

2.2.1 Media Spaces with Proxemic Cues. Prior work on media spaces has proposed to design office spaces and their communication hardware to provide social and spatial cues and investigate the utility of proxemic cues such as gaze direction and bodily orientation [65, 66], and/or apply spatial (proxemic) cues to connect the context of office space layouts in the workplace [13]. While these did not explicitly articulate their rationales in terms of proxemics, later works started to use proxemics to frame remote meeting interactions [57, 67], such as the media space of spatially aligning remote spaces (e.g., [55, 57]) and granting blended spatial consistency (e.g., [57]). Several systems have explored the gaze cues for mediating floor control in video communication [58] or visualizations [36, 75, 79].

In previous research, various methods have been proposed to compensate for the absence of real proxemic cues in distributed collaborations. [32, 37, 63, 72]. For example, Montage [71, 72] provided teleproximity for distributed groups as a mutual approaching glance fades in on others’ virtual workstations, enabling individuals to peek into each other’s offices. VideoWindow [20] designed “artificial proximity” to maintain informal interaction by directly translating physical proximity to a virtual environment. MirrorSpace [63] constructed proximity as an interface to provide smooth transitions between peripheral awareness and intimate forms of communication. PêLe-MêLe [32] supports different degrees of engagement by gradually shrinking and drifting videos towards the center of

the screen. FluidMeet [37] operationalizes proxemics for supporting private messaging and calls during multiparty video meetings by enabling different levels of interaction with others based on nested virtual interpersonal distances. However, this line of work utilizes proxemic cues for mediating interpersonal communication rather than individual or group conversation dynamics that include speakers, addressees, and listeners.

2.2.2 Virtual 3D Spaces with Proxemic Cues. Researchers have also explored the use of VR systems for mediating remote social interactions as the spatiality of VR allows more natural 3D proxemic relations [10, 12, 25, 78, 80]. Social VR provides opportunities for creating unique virtualized proxemic relations by manipulating properties of interpersonal relations such as the relative scale of avatars [80]. A few works have studied the relationship between turn-taking and embodiment in VR via conversation analysis [11, 12]. We share their interest in understanding how embodiment can be used to manage turn-taking; albeit focusing on the manipulation of video feeds instead of 3D avatars.

2.2.3 Virtual 2D Spaces with Proxemic Cues. A shared 2D virtual frame of reference can be used to establish proxemic cues without custom hardware setup. Some systems enable the blending of multiple video feeds to create a shared hybrid space [31, 52]. MirrorBlender [31] supports continuous repositioning, resizing, and blending of video feeds in a shared 2D interface using the principle of WYSIWIS (What You See Is What I See [68]). Unlike MirrorBlender, which supports blending of physical spaces in hybrid meetings, we focus on support for turn-taking in fully virtual meetings and more dynamic manipulation of person and task space. Most relevant to our work are the tools that support the free manipulation of video feeds, avatars, or UI elements to grab others’ attention in a shared 2D frame of reference [1, 5–7, 62]. Several recent videoconferencing (VC) tools enable similar experiences, such as proximity-based social interactions with avatars to trigger bubbles of conventional video-conferencing (e.g., Gather Town [1], Wonder.Me [7]), and/or repositionable video feeds to provide social awareness of parallel conversations with the proximity of everyone’s video with ad-hoc subgroup conversations (e.g., SpatialChat [5], Sprout [6]). Most of these interfaces provide some notion of virtual fixed/semifixed features, e.g., abstract rectangles or depictions of chairs or roundtables, that then provide a bounded space for scoping the video and audio channels to subgroups of currently online users (e.g., Remo [4]). The role of fixed/semifixed features in these interfaces is to support approach and leave-taking in groups [24] or larger online crowds [44]. Instead, our goal is to provide such spatial features for mediating *turn-taking* within groups that have already been formed.

2.3 Conversational Floor Management

A large body of research has investigated the role of non-verbal communication for meeting participants to take, yield, and maintain the conversational floor, e.g., [58, 66, 77]. The *conversational floor* can be defined as “a cognitively shared attentive space that mediates in the sequential or simultaneous organization of participants’ contributions” that users take turns and manage topics in a conversation [58].

2.3.1 Turn-Taking. Turn-taking describes the flow of participation among speakers in a conversation over time. It is through taking turns on the conversational floor that humans engage in dialogic collaborative problem-solving. The conversational floor is an “*dynamic and socially negotiated space*” where individuals have the opportunity to make contributions [17, 19]. Hence, examining the turn-taking structure is essential for understanding the interplay of various voices in a group. We are interested in how speaking turns shuffles among speakers proposed by the analytical tool called the participation-shift framework [21, 22]. The p-shift framework [21, 22], based on the work of Goffman [23], assigns the roles of the speaker, target (addressee), and third party (unaddressed recipient) to the participants in a conversation. The framework introduced four participation shifts¹ that accounts for possible micro turn-taking patterns, thereby illustrating the role of different participants in shaping the conversation and how turns are transferred from one speaker to another (Figure 2).

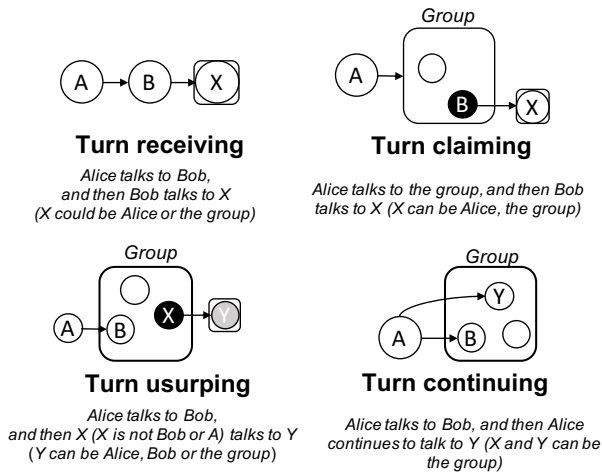


Figure 2: Four types of participation shifts as defined by Gibson [21, 22].

2.3.2 HCI Research on Turn-taking in Multiparty Meetings. In HCI research, most explorations into distributed collaboration employing videoconferencing (VC) or shared media space have primarily focused on pairs [38–40, 47, 59]. Research has also delved into how the medium impacts the user and conversational behaviors [16, 66]. However, with larger group sizes, meetings tend to be more moderated [26] and dominated by a few individuals [73]. Prior studies quantified turn-taking behaviors to compare face-to-face interactions with videoconferencing, examining various factors such as spoken characteristics (e.g., backchannels and overlaps) [56],

¹Four types of P-shifts: 1) *Turn-receiving* is described as the moment when an auditor takes over the speaking floor from the current speaker. 2) *Turn-claiming* happens when a speaker addresses the whole group and a third party responds to this invitation; 3) *Turn-usurping* happens when a third party who takes over (usurps) the speaking floor from the intended speaker (assigned by the current active speaker). This type of turn-taking can create disorder in group interactions; 4) *Turn-continuing* refers to the situation when a speaker retains control of the speaking floor while engaging in conversation with different individuals. This type of turn-taking can indicate a level of control or leadership in the group.

the influence of visual information [66], the provision of spatial cues [13, 65, 66], and head-turning and gaze cues [14]. A prominent example is Sellen’s study, which compared three VC systems with different levels of visual information shown to the users and revealed significant differences in how floor control and simultaneous talk were handled, particularly in comparison to same-room conversations [66]. For example, the study showed that same-room conversations exhibited a greater frequency of interruptions and fewer formal floor.

These findings remain relevant today; the conventional user interfaces for video-conferencing, i.e., Gallery View and Speaker View in Zoom [8], remain largely similar to those in Sellen’s study in 1995. However, with the recent emergence of 2D virtual spaces (with interpersonal proximity features) such as Gather [1] and Teamflow [9], it is timely to investigate whether such mechanisms can support remote users in handling who is speaking and/or who is being addressed. Our work aims to explore turn-taking patterns during video meetings to understand which participation-shift patterns [21, 22] emerge with OpenMic.

2.4 Conversational Transitions

Several studies of co-located collaboration show that social interactions are dynamic in their spatial nature. Lee et al. [46] analyzed how workers interact with each other in the workplace and found that there are many factors, other than the proximity between people, that affect socio-spatial formations and that socio-spatial formations change over time for various reasons, such as ergonomic and social comfort. Grønbaek et al. [30] found frequent transitions between different socio-spatial formations and demonstrated how shape-changing furniture might support these transitions. However, conventional videoconferencing tools with fixed video feed layouts and ordering, such as Zoom, make it challenging to achieve these conversational transitions. Emerging tools such as Gather, Wonder, and Teamflow have utilized proxemic metaphors for more organic transitions to support conversations. These tools enable Conversational Transitions (CTs) - as described by Gonzalez Dias et al. [24] - to support approaching and leave-taking. Inspired by this work, we propose the concept of **Conversational Floor Transitions**, which allows users to use proxemic metaphors to negotiate transitions non-verbally in video-mediated conversations.

3 OPENMIC: SUPPORTING CONVERSATIONAL FLOOR TRANSITIONS

We developed a prototype video-conferencing system, OpenMic, to explore how the use of proxemic cues may support turn-taking in multiparty remote meetings. To study small- to medium-sized group meetings (4 to 8 participants), OpenMic is built to scale beyond peer-to-peer. By routing the connections through an SFU², OpenMic can support more than eight video participants. The front-end and back-end communicate via WebSockets (i.e., a signaling server) to transmit data and HTTPS requests (i.e., authentication server) to access the APIs.

²Selective Forwarding Unit (SFU). OpenMic uses Ion-SFU, an open-source SFU server system that can be called directly or through a gRPC or json-rpc interface: <https://github.com/pion/ion-sfu>.

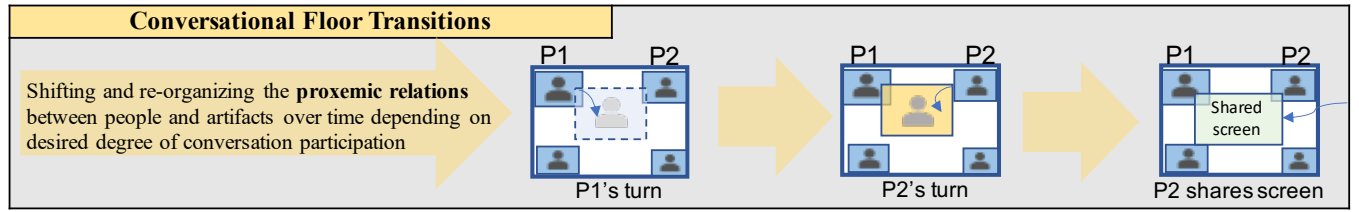


Figure 3: Conversational Floor Transitions (CFTs): Continuous UI transitions where users’ video feeds can change in relative size and position in relation to the conversational floor.

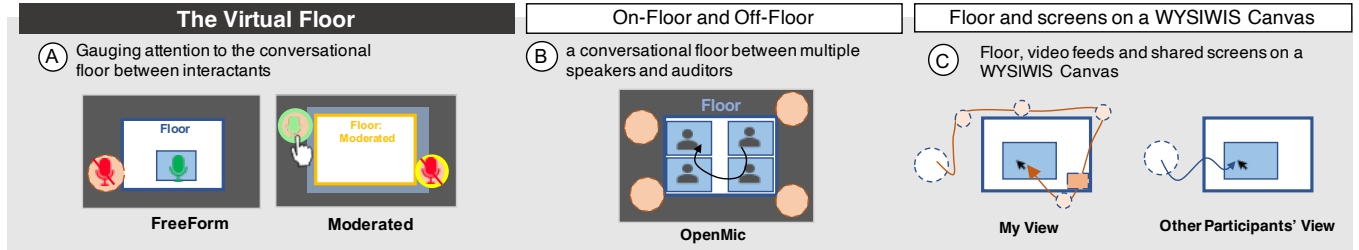


Figure 4: The Virtual Floor: A) Shared locus of attention regarding the *conversational floor* between speakers; B) Users appear as rectangles on the floor and circles off the floor; C) Rehearsing the Video feed manipulation in relation to the floor on a 2D WYSIWIS canvas.

The concept of Conversational Floor Transitions (CFTs) combines two proxemic metaphors: 1) The *Virtual Floor* which serves as a Fixed-feature space (3.1); and 2) *Malleable Mirrors* which allow for shifting the perceived proxemic distance (3.2).

When users move their videos (Figure 3), they have fine-grained control of their relative size and proximity to others for enacting their desired degree of participation. Continuous transitions on and off the Virtual Floor further allow for enacting different kinds of conversational floor transitions, e.g., addressing the whole group or single individuals.

In the following, we explain the design details of OpenMic regarding the interface concepts, Virtual Floor and Malleable Mirrors, outlining how they together support new communication means for conversational floor management, i.e., *Conversational Floor Transitions*.

3.1 The Virtual Floor as a Fixed Feature

In conventional video-conferencing systems, users use a dedicated button to mute and unmute for turn-taking, which often breaks the flow of conversation. With OpenMic, we aim to support a mechanism for non-verbally conveying intention to talk, which also implicitly mutes and unmutes. To help participants get a sense of the locus of attention in multiparty video meetings, we propose the concept of a *Virtual Floor* (Figure 4A) on a 2D What-You-See-Is-What-I-See (WYSIWIS) canvas (Figure 4C). For the purpose of readability, we refer distinctly to the interface concept as the *Virtual Floor*, whereas *conversational floor* refers to the concept of managing turn-taking in a conversation [58].

Moreover, the display layout follows the focus-plus-context approach [33, 41, 76]: the screen shows both an overview of the current attendees and a centered view of speakers that are at the locus

of attention in the meeting. The layout is shared among users on a strict WYSIWIS basis to help users relate one to another and support proxemic cues. Users can drag their video feeds around the WYSIWIS canvas (subsection 3.1C) to quickly switch between different roles, e.g., speakers, floor holders, and auditors.

Inspired by proxemics, the Virtual Floor is based on the idea of fixed/semifixed features [35], namely that humans use the furnishing of a room to enact their respective roles in human-human interaction. For instance, when people take specific seats around a table, they take on different spatial roles, where some positions signify intention to talk and others do not [45]. Another example is in conference theatres, where the furnishing affords focused attention on the stage area [45], and taking turns require the effort of moving across the boundary between the audience area and the stage where the performer is (or alternatively being passed a wireless microphone). Our Virtual Floor design most directly resembles fixed-feature space, as it is immovable in the interface. We discuss this shortcoming in Section 5.1. The following elaborates on the design properties of the Virtual Floor.

- **Design Property 1: Microphone Control.** OpenMic provides a *conversational floor* [58] as a fixed feature of the virtual 2D space, to which participants can respond like a piece of virtual “furniture.” Participants can use its edge as a boundary to draw attention to themselves while simultaneously managing their microphone, as the edges further define the boundary for muting and unmuting the microphone (Figure 4A). The term *conversational floor* is intended to be considered metaphorically, i.e., its purpose is to be appropriated by the meeting participants given their current turn-taking needs (e.g., current activity, group size,

formal/informal). In this paper, we study the role of the *conversational floor* and investigate how meeting participants give social meaning to the floor in the interface.

- Design Property 2: Configuring the Floor Boundary for Moderation.** Beyond simply allocating space for a conversational floor, OpenMic supports configuration of the floor area's role regarding turn-taking: the degree of moderation can be configured. The floor accommodates for the fact that increasing group size for meetings requires more moderation [26, 45] via two different modes of microphone management: *freeform conversation* and *moderated conversation*. Freeform mode (Figure 4A Freeform Mode) has an open boundary where users can move freely across the edges of the floor to mute and unmute. In Moderated Mode (Figure 4A Moderated Mode), the floor edges serve as a closed boundary around the floor. The moderation role is given based on the participant's location; all users on the Virtual floor become moderators. As the floor has a closed boundary, anyone who wants to be on the floor needs to be approved by a moderator. Unlike other video conferencing tools that have designated hosts or moderators, OpenMic utilizes the spatial relationship to the fixed feature, i.e., the Virtual Floor, to allow fluid switching of the roles during conversations. For example, when there is a series of small group presentations, any member of the group on the floor can easily hand over the floor to the next group or an individual with a temporary question. The speakers on the floor all have controls for moderating whether individuals from the audience can enter. When audience members bring their video feed to the boundary, there will be a yellow ring visualized around the audience's video feed. As everyone (speakers and audiences) can see the yellow ring color and the video feed adjacent to the boundary (Figure 5A-B), speakers can click their video feeds to grant access to speaking and moving to the floor. The yellow ring around the video feed will then become a green ring (with the red muted icon changed to green unmuted specifically for the granted audience) so that everyone in the meeting can notice the audience being granted to speak (See Figure 5B-C).

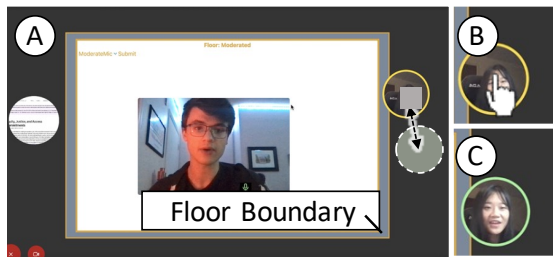


Figure 5: Floor Boundary

3.2 Malleable Mirrors On and Around the Virtual Floor

Interaction techniques with video feeds and shared screens are similar to Grønbaek et al.'s principle of *Malleable Mirrors* [31]. *Mirrors* are video windows with mirrored images of people or streams from

screen sharing. Making mirrors malleable means to enable manipulation of their properties. We extend this principle by binding these properties (Design Property 3). When physically co-located, we rely on the interpersonal space and are able to manage turn-taking in conversations naturally by relying on conventions from cues such as bodily orientation and mutual gaze [35, 42]. Moreover, E.T. Hall [35] outlines how our physical distance to other people correlates with our ability to perceive aspects of them. At public distance, we mostly perceive large bodily gestures, whereas at social distance, we can clearly see facial expressions and nuanced hand gestures, and at personal distance, we can perceive detailed eye movements and gaze direction. In video-conferencing, these proxemic relations do not disappear but instead, they are perceived virtually, where increased video size becomes a signifier of decreased proxemic distance; this is termed Perceived Proxemic Distance [18, 27].



Figure 6: Malleable and Projected Screens: A) Resizable Screens. B) the user can click on the “project” icon to project full screen to the Floor;

- Design Property 3: Mapping Scale and Position of Malleable Mirrors.** When interacting with Malleable Mirrors within the Virtual Floor boundary, we apply *simultaneous transformation* of position and scale. This technique is inspired by the concept of Perceived Proxemic Distance. Hence, in OpenMic, video position and size are integrated in a single mouse motion so that the resizing and re-positioning occurs in parallel rather than sequentially (Figure 7B). Upon entering the floor boundary, the relative position of the cursor to the center is mapped to video size, meaning that the video size increases gradually as the participant moves the cursor closer to the center of the floor. The video size will expand to its maximum near Floor Center (Figure 7C-1). After the video feed reaches its full size (Figure 7C-1), repositioning the video feed will not result in any video size changes (Figure 7C: 1-2), as long as the video feed boundary is within the Floor boundary (Figure 7D-1). The user can gradually shrink the video in reverse as the video feed boundary reaches the Floor boundary (Figure 7D-2). The unified continuous mapping of position and size allows for the participant to implicitly reduce how others perceive the proxemic distance to them, as they move towards the center of attention of the conversational floor. This continuous positioning and resizing of video feeds also applies to shared screens (Figure 6A). In addition to moving their shared screen to the Floor area, participants can project the intended screen to the floor by double-clicking it (Figure 6A-B). The continuous control of the degree to which one is in the perceived center of attention enables participants to *gradually* enter and leave the floor. Although the size of a person in the video framing can vary

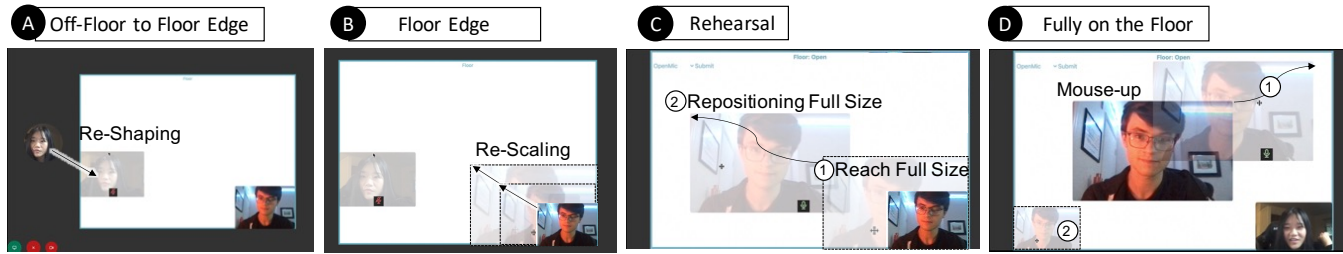


Figure 7: Malleable Mirrors (Freeform Mode): A) the video feed will be reshaped during the transition from off-Floor to Floor Edge; B) when entering the floor boundary, repositioning the video feed will continuously resize it until C-1) it reaches its full size; C-2) the user can then reposition the video feed within the on-Floor area without video size changes. D-1) The video feed will stay full size until D-2) it touches the Floor boundary, indicating a quit-Floor activity. The video size will thereafter decrease.

as the person leans closer to or away from the camera, OpenMic gives virtual control of a person’s apparent size.

- **Design Property 4: Shape of Malleable Mirrors.** In the transition of taking the floor, the shape of the video feed changes from a circle to a rectangular shape that shows the full video feed. The *circular crop* is intended to maximize the view of the participant’s face (for conveying head nodding) while minimizing the space taken up in the periphery of the floor area. The *rectangular shape* is intended to show the space around the participant’s head for conveying hand gestures. Figure 7A shows the transition of how participants outside the floor area take the floor and reshape when entering the floor.

A final extension to Malleable Mirrors is that users can rehearse their interactions without being observed by others. The ability to rehearse interactions in a personal space has been shown to be important in shared virtual spaces [34, 69, 70]. Users can drag their video or screen feeds, seeing a temporarily semi-transparent version of the video feed on the 2D WYSIWIS canvas to rehearse their potential configuration before sharing the path before the mouse-up event. With the mouse-up event, other participants can see the semi-transparent feed animate to the position of the cursor along the shortest path from its original position (Figure 4C).

4 EXPLORATORY STUDY

We conducted an exploratory study with OpenMic to gain insights into how users use the Virtual Floor and video feed manipulation to interact with each other and how these features enable novel turn-taking behaviors in video meetings. To qualitatively assess the effect of OpenMic, we used Gibson’s participation-shift framework [21] and analyzed how the proxemic cues through conversational floor transitions are used to invoke different types of participation shifts. We designed a series of tasks to specifically invoke such dynamics and conducted a qualitative observational study consisting of two multiparty meeting sessions.

4.1 Participants

We recruited 16 participants (8 females and 8 males) from the local university (average age = 24.3). The participants were split into two groups. Their familiarity with videoconferencing tools was high ($M=6.28$, $SD=0.19$, measured on a 7-point Likert scale from 1 to 7). All participants have used Zoom. Six of them have used Microsoft

Teams. Participants were provided with \$20 as compensation for the one and a half hour experiment.

4.2 Tasks and Procedure

Upon arrival to the virtual meeting room, participants filled out a consent form and initial demographic survey. After this, the researchers explained the purpose of the study to the participants. We asked one remote participant to volunteer as a meeting moderator and team leader in each group. Each study session lasted for about an hour and 30 minutes and consisted of a training session, three tasks, and an interview. Throughout the study, the experimenter was in the virtual room of OpenMic, without interrupting an ongoing task. Participants made their own decisions about how to use OpenMic features, e.g., when to use FreeForm or Moderated Mode. When one task ended, the experimenter would advance to introduce the next task of the study. In some tasks during the session, the group was divided into breakout rooms (separate running instances of OpenMic) in four-person subgroups, before reconvening in the same room.

4.2.1 Pre-Study Training (20 mins, eight-person group in one room).

Before the study, the research team demonstrated how to use OpenMic, including moving video feeds and sharing screens, and how OpenMic’s Moderated and Freeform mode works to users. This was done via video calls with a live remote demonstration. Participants were then asked to try OpenMic on their web browser and were encouraged to freely explore the interface to get familiar with OpenMic.

4.2.2 Turn Taking Task (5 mins, eight-person group in one room).

The first task was to introduce themselves to the group. After a brief introduction, the eight participants were split into two subgroups of four in two breakout rooms.

4.2.3 Ice-Breaking Task (15 mins, in four-person groups).

After joining the breakout room, the participants started an icebreaker game called “One truth and two lies.” The moderator instructed each participant to think of three statements about themselves: two must be true statements, and one must be false. Others, as audiences, could ask questions about the statements. After three rounds of questions, they discussed and decided which one was the lie. We chose this task to resemble a meeting where one speaker controls one’s presence on the Virtual Floor, regarded as the floor holder and others are supporting through audience participation and other



Figure 8: A new presenter group enters the Virtual Floor (in an eight-person group). A-B: A group is moving themselves to the floor (before switching to Moderated Mode). B-C: Transition of group presentation and switching groups from one group to another. C: The presenting group members from the second group reserved themselves to the edge of the floor first, and then adjusted their scale relative to others.

short interjections. We particularly observed dynamic use of video position and scale in the conversation.

4.2.4 Survival Task (50 mins). Participants will be assigned to discussion sessions with a popular team-building task called Lost at Sea [2]. We assigned a Google sheet with survival items to each group and provided the group and individual sub-sheets. After receiving their scenario, participants were given up to ten minutes to rank their items individually in their individual sub-sheet in the order of most helpful to least useful for survival.

- **Part 1 Breakout Task (25-30 min, in four-person groups):** Eight participants stayed in the same two four-person groups as the ice-breaking task, and each group decided the top items of four in a group discussion with a maximum of 30 minutes. We chose this task to resemble a team discussion and decision-making meeting where participants would discuss their opinions. The task also required them to reach a consensus. We were interested in the transitions between different virtual workspace arrangements of people and screens.
- **Part 2 Group Task (15-20 min, eight-person groups):** After reaching consensus, the two groups gathered in one room to present their rationales by groups. They decided which group presented first. When one group was presenting their choices and rationales, another group remained off the floor as the audience. After each presentation, there was a Q&A session where another group could challenge the speaker group. We chose this task to mimic group presentations where multiple speakers are on the floor and audiences pose questions to the speakers. We were interested in the behaviors on and off-Floor, and potential moderation of speaker-audience conversations.

4.3 Data Collection and Analysis

We collected two types of data during the study. All data collection and analysis for this study was approved by the IRB.

4.3.1 Video Data. The study sessions were screen recorded. During the study, two researchers collected field notes to capture interesting episodes of turn-taking behaviors. We observed the participants' behaviors, such as how they communicated non-verbally through the video movement (via re-positioning and re-scaling) and how they approached or retracted from the Virtual Floor for participation shifts.

We conducted a top-down (deductive) thematic analysis on all the notes. We first discussed preliminary themes that aligned with the three proxemic metaphors (*fixed/semifixed features*, *perceived proxemic distance*, and *conversational transitions*) and then grouped the notes into themes related to these. When revisiting the video data, we focused on episodes (see Figures 9–17) that are specifically related to these themes (see themes below). Two researchers (the ones who collected the field notes) were involved in the thematic analysis. They also cross-referenced the focus-group findings with the key episodes from the video analysis in order to re-review and further refine the notes and key events from the recorded video with the three main theoretical themes. The findings were finally discussed and organized by all researchers on the team.

4.3.2 Focus Group. To complement the video data and field notes with a qualitative understanding of the users' experiences, focus group interviews were conducted with all participants in their four-person subgroups after the three study tasks. This resulted in four focus group sessions. Each focus group interview session lasted about 20 minutes. The interview probed participants general use of OpenMic features. We specifically asked the groups about interesting and surprising episodes and field notes that we observed during the study tasks. The focus group interviews were screen and transcribed.

4.4 Findings

This study showed diverse use of the system that connects socio-spatial theory to turn-taking behaviors. The findings are structured in the following three themes: *Use of the Virtual Floor* (4.5), *Turn-Taking Behaviors* (4.6), and *Re-configuring for Different Conversation Styles* (4.7). Under each theme, we report on a set of vignettes to describe the details of interesting episodes of turn-taking behaviour.

4.5 Use of the Virtual Floor

The first group of observations regards how the Virtual Floor serves as a fixed-feature space that bounds the shared attention, enabling participants to manage the conversations in the meeting.

4.5.1 Perceived Personal Space Is Different On vs. Off the Virtual Floor. Our observations show that the Virtual Floor provides a visual separation of the 2D virtual space that differentiates how

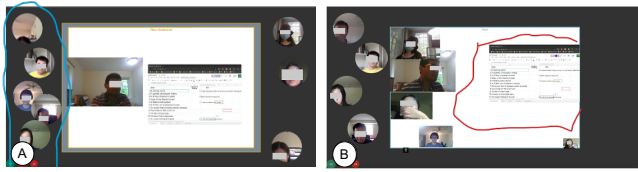


Figure 9: Personal space off and on the Virtual Floor (in an eight-person group). A: When participants entered the presentation room, they were crowded near the entrance. B: One four-person group moved to the Virtual Floor to deliver a group presentation; on the Virtual Floor, they were aware of avoiding overlapping videos.

participants behave and perceive each other *on* and *off* the Virtual Floor in terms of their perceived personal space.

On the Virtual Floor, we saw indications that participants embody their video feeds as their *personal space*, which is consistent with the findings of Grønbaek et al. [31]. For example, during the switch to a new presenter group (Figure 8), participants on the Virtual Floor tended to move closer to the boundary to leave space for others. The focus group interview confirmed this analysis of how participants interpreted the Floor as an open space and indicated respect for the personal space of one another: “I consciously thought about open space for more people, and we do not need to cover all the space, we may leave some spaces for others to join.” “...resemble personal space, in the real world, would you stand on top of someone? No, you wouldn’t.”

Off the Virtual Floor, however, we observed that participants tended to stay at the initial position of where they appear upon entrance with two or three participants overlapped with each other (Figure 9A:top-left). They did not move their video feeds unless they intended to take turns (Figure 9B).

This observation is corroborated by the focus group interview, when one participant recommended supporting a transition from the structured state of “*taking seats in the dimmed area or around it [the periphery of the floor]*” to the unstructured state in OpenMic so that participants perceived that they can benefit from “*whenever you want to take a turn, you can enter the floor and re-sizing our video.*”

This observation further indicates that the Virtual Floor provides a clear visual separation to indicate where the shared attention is. Because the floor area in the presentation scenario of Figure 8 and Figure 9 serves the function of a *stage* for groups to present in front of others, personal space becomes important to have clear views of each presenter, but less so for the audience, who is not the center of attention.

4.5.2 Use of the Floor Boundary Depends on the Group Size. We found the differences in the way participants enter and leave the Virtual Floor by the group size, particularly in how the boundary mediated their turn-taking behaviors. Even when the breakout groups used a Moderated Mode, they usually kept themselves on the boundary and maintained the conversation flow for a long duration. For instance, Figure 10A illustrates an example in a four-person group when three participants kept maintaining their position on the boundary for several minutes after their audio got approved

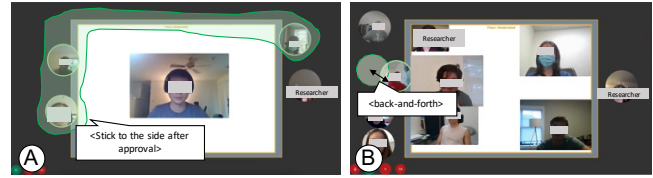


Figure 10: Group size affects the use of the Virtual Floor (in Moderated Mode). A: In a four-person group, the participants did not move away from the boundary and kept their microphones on; B: In the eight-person group, the *Off-Floor* participants entered and left the Virtual Floor more frequently.

by the speaker. In contrast, Figure 10B shows an example that in a larger eight-person group, a participant near the moderated boundary kept entering and leaving the Floor four times to actively engage in the conversation. During the focus group interview, a participant mentioned how they interpreted the space on and around the Virtual Floor and used it for turn-taking in group conversations: “I thought [that] when you take the floor [it is] like when you stand up in the classroom.” To further illustrate, the participant described his experience in the classroom that he would sit back down when he did not contribute to the conversation.

Group size also influences the moderation of the Virtual Floor when there are multiple moderators on the Virtual Floor. When there was one speaker on the Floor (typically during the ice-breaking task), it was simple for that speaker to control the crowd and provide permission to speak without any problems (Figure 10A). However, we discovered some difficulties in how multiple speakers managed the Virtual Floor and coordinated audience permission to speak. During the eight-person group with four speakers on the Floor, one speaker asked the other on the floor to unmute participants in the audience. When two participants in the audience approached the boundary with the intention to speak, they were frequently allowed (shown by a green ring surrounding the video feed) permission to speak at the same time. This caused confusion and one from the audience would have to withdraw their video. During the focus group interview, the participants confirmed the confusion over who grants participants off of the floor access.

4.6 Turn-Taking Behaviors

Resizable and repositionable video feeds enabled participants to use their size, position, and distance to others to indicate their turn-taking intentions. Moreover, the reconfiguration and the Virtual Floor’s boundary supported the diverse micro-modification of video movements.

4.6.1 Turn-taking Behavior during Presentations Exploits the Virtual Floor as a Stage. On-Floor participants made consistent use of their malleable video feeds for taking the turn - both re-sizing and re-positioning on and around the Virtual Floor for explicit communication with other members.

The following two episodes illustrate two different ways of taking the turn during the group’s presentation: (1) moving to the center area of the Virtual Floor; (2) slightly enlarging themselves

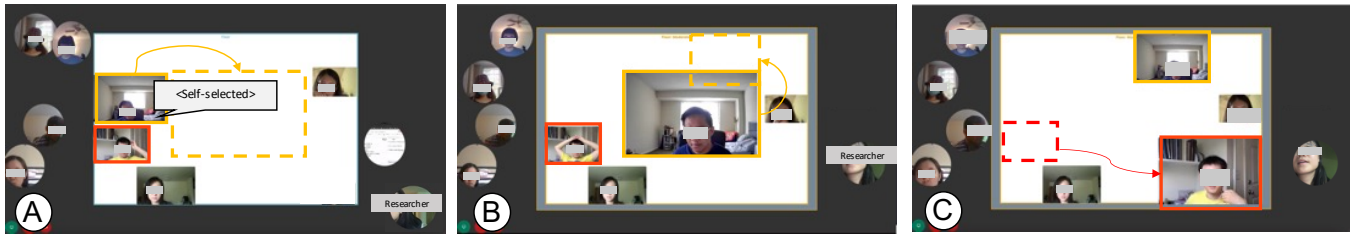


Figure 11: Screenshots for Vignette 1 (Eight-Person Group, Presentation, Survival Task): Taking turns while on the floor (dashed shapes indicate the previous or next video positions).

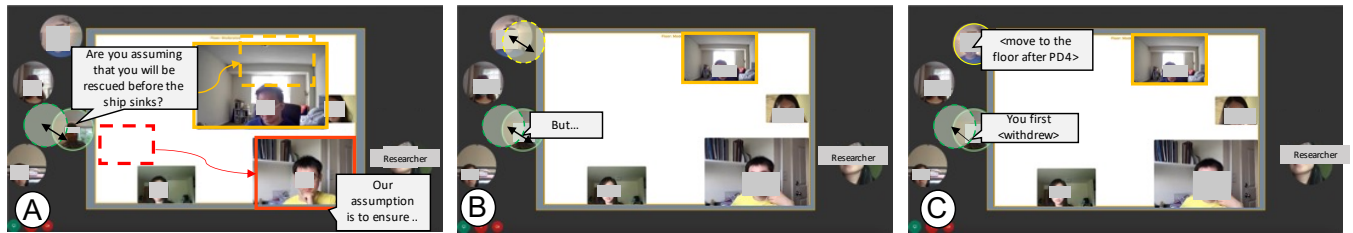


Figure 12: Screenshots for Vignette 2 (Eight-Person Group, Q&A session, Survival Task): A transition from presentation to Q&A.

near the edge by moving closer to the floor center. Vignette 1 (Figure 11) shows an episode where a participant shifted the "stage" of Virtual Floor by first putting himself as the next speaker by moving his video feed to the center of the Floor, and he moved to top peripheral area after finishing his turn (Figure 11C):

Vignette 1 (Figure 11): Eight-Person Group, Presentation. Group 1 entered the Floor to start their group presentation, and each of them placed them at the edge of the Floor. There was a short silence before anyone wants to start (A). P1 (yellow) self-selected as the next turn-taker. He started the Moderated Mode and moved his video to the Floor center to present his group's idea (B); After P1 finished his presentation, he moved himself to the upper-right position (C).

Another way can be illustrated in Vignette 2 (Figure 12) that a participant enlarged himself a bit to answer the question by an audience member during the Q&A session:

Vignette 2 (Figure 12): Eight-Person Group, Q&A session. There was an audience question by P2 during the Q&A (A); P2 withdrew his video after asking the question. P3 from the presenter group (red) moved himself to the side (and then enlarged himself a bit) to answer the question (A); P2, as the audience, moved himself to the boundary (approved by speakers), saying "But...", while he saw the intention of another audience (P4) with yellow rings around his video feed (B-C). P2 then said "you first" and withdrew (C).

These episodes show how *on-Floor* participants use the size and position of their video feeds by shifting between the center and peripheral area with their concurrent "self-selection" mechanism. The Virtual Floor as a stage to configure multiple co-speakers and align and re-configure their malleable video positions was also of significant utility to participants to enhance participation and grab attention from both other speakers and *off-Floor* audiences.

4.6.2 *Relative Position between Videos on and around the Virtual Floor.* While resizable and repositionable video feeds enabled *on-Floor* participants to grab attention, *off-Floor* participants made consistent use of their relative position to communicate with different subsets of the group, e.g., address a group, a specific speaker and maintain awareness of other audiences' behavior.

Often, *off-Floor* participants took the floor using the shortest path to the Virtual Floor (e.g., Figure 13A). This shows a *turn-claiming* pattern in which audiences reversely address the whole presenter group via the boundary of the Virtual Floor, and one or two of the co-presenters would respond to the question (e.g., Figure 10). We also observed some cases in which *off-Floor* participants move to a specific position of the boundary to have a conversation in relation to someone (i.e., address a specific person) on the Floor. This shows a *turn-usurping* pattern when one of them takes the floor and addresses a specific speaker to continue or adds a point of another turn-taker. An example was illustrated in Vignette 3 (Figure 13) where one audience member (P3) prominently moved her video feed right next to the boundary near the current speaker (P1) to take the turn of what P1 addressed to another audience member (P2), and built up on the conversational threads between them (P1 and P2):

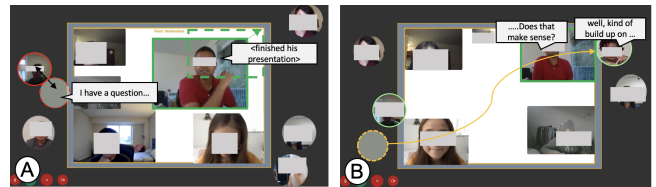


Figure 13: Screenshots for Vignette 3 (Eight-Person Group, Q&A Session): Turn-usurping indicated with video changes.

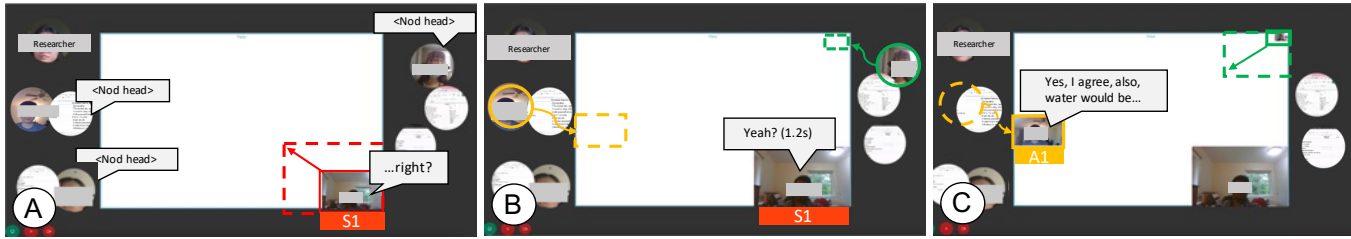


Figure 14: Screenshots for Vignette 4 (Four-Person Group, Survival Task Part 1): Conveying intention to hand over the floor.

Vignette 3 (Figure 13): Eight-Person Group, Q&A session. Right after the speaker (S1 in green) finished his presentation, he moved himself to the corner of the Virtual Floor (top-right); the auditor (A1 in red) drew his video feed to the moderated boundary and asked a question (A); S1 then responded to him ending up with “Does it make sense?”. Another audience (A2 in yellow) moved herself from the bottom-left to top-right (farther path but closer relative position), saying “well, an adjacent question to [A1]’s question...” (B).

The finding shows how the *relative position* around the Virtual Floor and the proximity to other participants’ (especially a speaker) video feeds provide participants with more opportunities to express their intention to others. The relative position (being able to enact closer interpersonal distances) also aided communication as the user can point their video feeds to a specific individual as an indicator to address a person.

4.6.3 Resizing the Video Can Convey the Intent to Hand Over the Floor. As illustrated in prior sections, speakers (on the floor) and auditors (off the floor) made use of the re-positioning extensively for turn-taking and role switch between central and passive speakers in the **Survival Task** (eight-person group) for group presentation and Q&A tasks.

Here, we also observed two ways of how speakers used video resizing to convey intentions of *turn-claiming*³ in four-person groups. First, one episode (Figure 16B-C) shows how *decreasing video size* can indicate intent to give the floor to others: the participant slightly moved towards the corner to invite others to take the floor. Second, as an opposite example, the episode in Vignette 4 (Figure 14A) shows how the speaker subtly *enlarged the video* as a nonverbal form of conveying the intention to the whole group to reach an agreement. It then shows how this interaction invites other group members to enter the Virtual Floor to engage in the conversation (Figure 14B-C).

Vignette 4 (Figure 14): Four-Person Group, Group discussion. S1 (red) tried to hand over the floor verbally with “I guess if everyone agrees, we can put 1 and 4 and the seat cushion, right? That would work?” Other group members responded by nodding their heads (A). He then moved himself closer to the Floor center and hence subtly enlarged his video feed on the Floor, saying “yeah? Or I will just put it in the sheet” (A-B); Other group members started to enter the Virtual Floor and address their ideas. A1 responded “Yes, I agree; also, water would be the second choice, and for the third, we kind of divided on that.” (C).

³Turn-claiming happens when a speaker addresses the whole group, and a third party responds to the invitation.

During the focus group, participants reported mismatches between their understanding of the conversation state: “I nodded my head a lot during the tasks mostly because I thought that the speaker was gonna continue his statement and ideas so that I would not interrupt, [...] but later he enlarged himself, and it was obvious that it could be my turn to say something.”

One reason for this mismatch is that the request was interpreted differently by other group members as a confirmation rather than a handover request as participants further commented on the process of addressing the whole group is sometimes a hard-to-tell intention in video-mediated communication: “we can explicitly mention someone’s name, when you want to address a specific person. [...] When someone asked the whole group, it is often hard to tell when the turn goes next in the video meetings. Did the speaker finish his turn? [...] In this case, OpenMic did a great job.” Another possible explanation regarding this challenge of addressing the whole group has been found in prior work [66] that people are relatively polite in video meetings when taking the turn and less tenacious about holding the floor, which can be potentially alleviated by adjusting the size of video feeds to mediate their concurrent talk in OpenMic.

4.7 Re-configuring for Different Conversation Styles

The following shows episodes where groups transition between different spatial configurations of video and screen feeds. We analyze how they re-configure the 2D workspace for different conversation styles to achieve the desirable turn-taking properties for the given task.

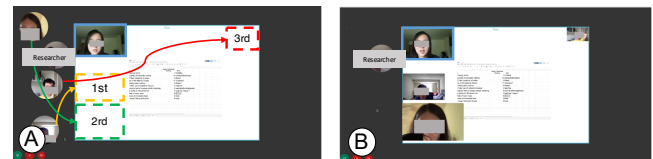


Figure 15: Turn-continuing with persistent video size and position (in a four-person group). A-B: P1 (in blue) was presenting her choice when other group members asked her questions by moving to the floor sequentially (labeled number as the order), but P1 was still leading the discussion.

4.7.1 Enhanced Awareness of Intention Helps Avoid Confusion. In the presentation configuration, speakers are on the Virtual Floor, whereas the audience is distributed around it, which supports workspace awareness during Q&A: crossing the boundary shows

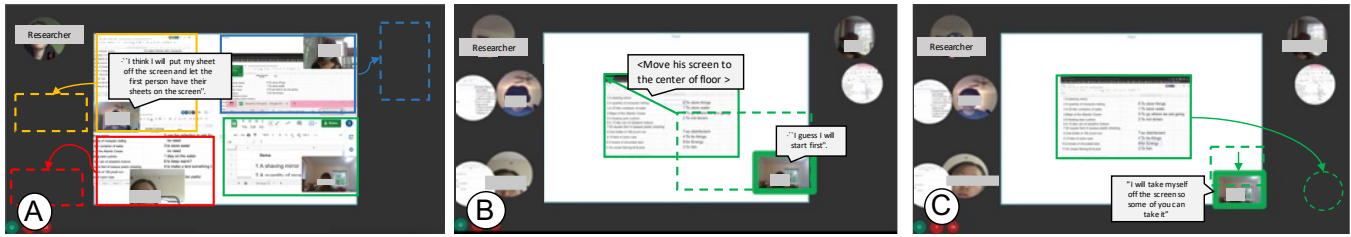


Figure 16: Screenshots for Vignette 5 (Four-Person group): Workspace transitions and Re-organization of Shared screens.

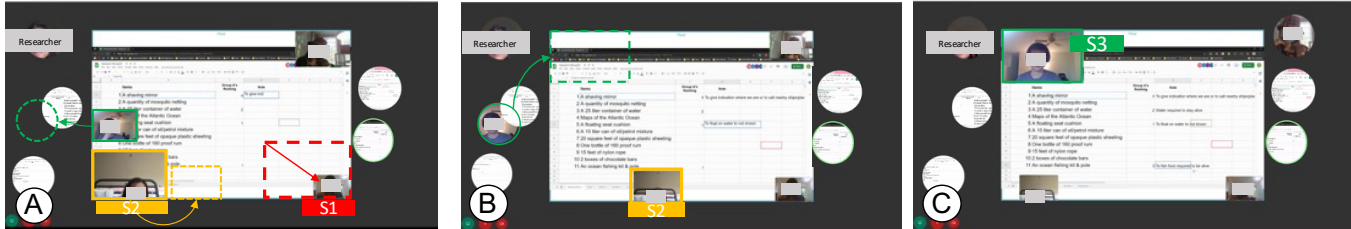


Figure 17: Screenshots for Vignette 6 (Four-Person Group): Re-organization of large shared screens.

clear intention to talk, which helped prevent potential confusion regarding who is next to speak. For example, Fig 12B-C illustrates such an example where the audience member withdrew his video feed as he saw the approach of another audience member to the boundary and understood his intention to take the floor.

4.7.2 The Organization of Shared Screens Shapes Turn-Taking Patterns. The support for multiple simultaneously shared screens distributed around the Virtual Floor helped mediate turn-taking. We will show two episodes in four-person groups that show interesting ways of using screens to handle turn-taking conflicts. For instance, Figure 15 details a workspace organization that shaped the given group's turn-taking patterns during artifacts sharing, where one participant maintained the central role of leading the discussion while other members momentarily jumped in to the conversation to question her choices (without showing their own list). However, Vignette 5 in Figure 16 shows an episode that represents a conversational transition from a configuration that enabled item-to-item discussion of comparing their individual selections around each other's shared sheets (Figure 16A) to a configuration supporting quick shifts between individuals presenting (Figure 16B-C).

Vignette 5 (Figure 16): Four-Person Group, Group discussion One participant was presenting his selection of items, followed by a quick round of discussion about the top item and comparison between each other's decision and sheets (all the participants moved their screens to the floor to point out their rationales over the item). They then found it a bit hard to come to an agreement. One participant proposed to go through it one by one (A); Participants then put off their screens. One participant (green) claimed the turn to be the first to speak and moved his screen to the center (B). After he finished his turn, he slightly moved his video feed towards the bottom-right corner, saying "I will take myself off the screen so some of you can take it" before he leave the floor (B,C).

The above episode shows that the re-positioning of screens prevents potential conflicts regarding who is taking the next turn to

speak. On the contrary, the current implementation of sharing screens in full-screen (*standout*) also caused issues regarding how to position person videos within the limited display space. The episode in Vignette 6 (Figure 17) shows that when screen-sharing needed to cover the entire Virtual Floor, users struggled with how to position themselves to avoid occluding the content of discussion.

Vignette 6 (Figure 17): Four-Person Group, Group discussion Once each member of a four-person group had presented their own ranking of the survival task, there was a three-minute discussion during which each group member was on the Floor, yet none of them shared screens or moved their video feed (A), and they viewed their dual screens and discussed their group's decision aloud. Finally, after a discussion about an item, S1 was trying to confirm the choice with his group members, when he looked back to the primary screen on his laptop: "Okay, then I will put the third one as the shaving mirror...wait, I will share my screen." After the screen was shared, the origin position of S2 and S3 appeared to cover the shared screen. They then moved themselves to other places (A,B). Later, S3 avoided covering the display by shifting to the upper-left edge (C).

The participants' need to stay on the floor is that it is what keeps them unmuted and thus technically allows them to stay engaged in the conversation. Hence, this vignette shows that the support for conversational floor transitions with person videos is competing with screen-sharing to use the limited screen real-estate on the Virtual Floor.

5 DISCUSSION

Here we outline the main implications of our work, based on the prospects and limitations raised in our studies. We further point to future directions for supporting turn-taking with interface concepts inspired by proxemics theory.

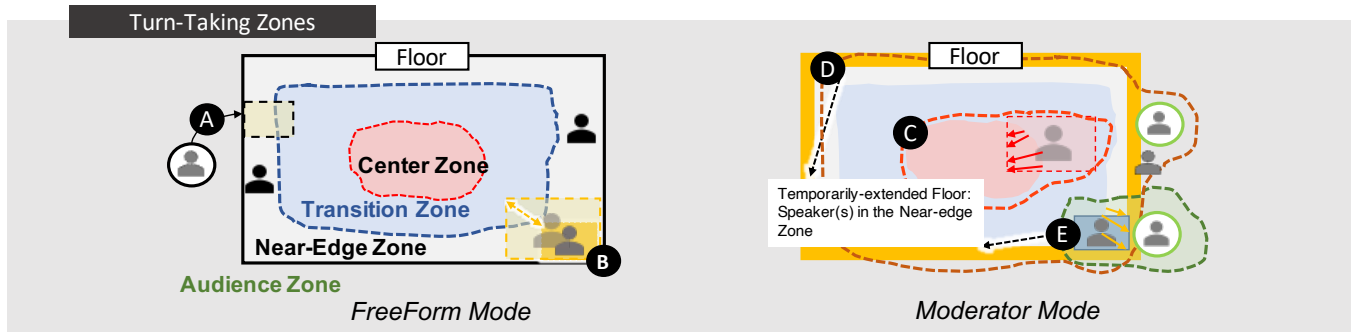


Figure 18: Turn-taking Zones: based on observed turn-taking patterns, we categorize users’ partitioning of the Virtual Floor area into four zones; *Audience, Near-Edge, Transition, and Center*.

5.1 Support a Variety in Group Sizes with Fixed/Semifixed Features

Proxemics research has found that an important factor determining how we organize in relation to fixed- and semifixed-feature spaces is group size [45]. In co-located small-group meetings, we tend to organize around roundtables (affords equal participation), whereas larger meetings may be held in lecture theatres (affording moderation) [45]. We saw many similar patterns in the social interactions emerging across our studies with OpenMic. Comparing full-group (8-person) and breakout-group (4-person) activities, it is clear that in small-group meetings (4 persons), participants tend to negotiate the Virtual Floor space akin to how we engage with a “meeting table” (i.e., staying seated, gesturing to take the floor) with relatively static patterns, whereas even in medium-sized meetings (8 persons), the dynamics change to resemble more formal social interaction patterns, akin to lecture halls (i.e., raising the hand, standing up to speak) with rich conversation dynamics and conversational floor transitions. This result also makes sense given common issues with microphone management in video meetings; in small-group conversation, participants rarely need to mute, but may exploit the area near the floor edge or video re-sizing for turn-taking cues. On the other hand, larger groups may appropriate the area as a “stage” with open mic, where participants can “walk on stage” (unmute), or with moderated mic control, where participants are given permission to “walk on stage”.

The implication for future interface designers interested in developing virtual fixed/semifixed features, is to support a variety of group sizes, with different levels of moderation. As group size increases, meetings tend to be more moderated [26], and the OpenMic design provides an example principle to support this variety, with a switch for moderating entry to the Virtual Floor (Moderator Mode vs. FreeForm Mode). The mode switch only scratches the surface of ways that fixed/semifixed features can be configurable. Our Virtual Floor design resembles more *fixed-feature space*, as the position and size of the Virtual Floor frame were fixed. However, it could be interesting to explore the impact of designing the Virtual Floor as a *semifixed-feature space*; e.g., enable changes in position and size of the Virtual Floor frame to cover different areas and amounts of the display space, or support the creation of multiple floor frames for breakout room conversations [37]. This way, participants could “refurnish” their virtual 2D space for turn-taking properties that are

desirable to the given group size and activity. Moreover, semifixed floors may support a larger variety of meetings, and may affect the turn-taking behavior (e.g., a smaller floor will provide a bottleneck for the number of current speakers that can appear at a reasonable size).

5.2 Turn-taking Zones: Trade-offs in Free vs. Curated Positioning of Videos

Analyzing the episodes of the exploratory study, we see patterns emerging in how participants the virtual 2D space, whether interacting in small- or medium-sized groups. We categorize users’ partitioning of the virtual space into four types of *Turn-taking Zones*: audience zone, near-edge zone, transition zone, and center zone (Figure 18). The progression of these zones resembles the concept of proxemic distance zones [35]. A similar HCI concept is seen for ubicomp applications with proxemic interactions [28] and gradual engagement [48]. In our case, these zones mediate the gradual engagement with other participants in distributed video meetings.

In both group presentation tasks and breakout-group tasks, 4-person groups took the Virtual Floor area. In these situations, the Near-Edge Zone was used as a temporary peripheral zone, reserved for their video feeds (in a smaller scale) closer to the boundary (Fig. 8A). Episodes of movement in the Transition Zone showed diverse and rich interactions with subtle proxemic cues (as outlined in Section 4.6). The movement across to the Center Zone shows the change of conversation state, and overall we identified three different strategies: 1) using the Near-Edge Zone to show the intention of taking the floor (waiting for others to finish the turn), shifting from an auditor to the speaker role; 2) moving to the Transition Zone directly to start a turn, speaking up in the discussion; 3) moving swiftly from the Near-Edge Zone to the Transition/Center Zone and back to the Near-Edge Zone again, to shift between being passive co-speaker and central speaker that is currently speaking.

The emergence of these zones is also consistent with studies of territoriality in proxemics literature [35], HCI studies of collaborative interfaces [64, 74], and 2D virtual interfaces that provide video embodiment [31], where people respect boundaries of personal space. In our case, this was seen in how meeting participants avoided overlapping video feeds, and in how they continually negotiated the real-estate of the 2D workspace in tasks crowded with screen feeds, e.g., the more successful cases of turn-taking were

occurring in the eight-person groups (vignettes 1-3), whereas for four-person groups, the use of the boundary was less frequent and the full-screen screen-sharing conflicted with the use of the floor space to stay engaged (unmuted) in the conversation (vignette 6).

This highlights a trade-off regarding fine-grained video movement vs. binary and coarse-grained control of video feeds. Comparing small-sized and medium-sized group dynamics, one conclusion might be that it had negligible value for small-group dynamics, whereas it would provide more value for larger groups, and perhaps even more so for the tasks that involve more formal negotiation of the Virtual Floor space, such as groups switching between taking the floor to share screens. Future designers of video-conferencing interfaces with proxemic metaphors may thus consider supporting a better balance between structured and unstructured layouts of video feeds. With more structure implied by having several fixed-feature elements, it may reduce the amount of manual effort, e.g., by enabling videos to snap to zones in the interface (akin to e.g., Sprout [6] or Teamflow [9]). On the contrary, more flexible movement of videos also provides more opportunities for expressing interpersonal relations virtually [31].

5.3 Support for Different Kinds of Proximity Cues

Our analysis of turn-taking events via virtual proxemic metaphors revealed compelling insights for how the OpenMic design may address the issue of users' confusion regarding others' intention to talk, thus supporting conversational floor transitions. Prior work found more explicit (or formal) handovers in video-mediated communication, where this formality came from addressing others by name, as video communication usually lacks gaze or other implicit cues [66]. Our evaluation of OpenMic indicates that movable and re-sizeable video feeds may substitute the verbal handover of turns by making the handover cues more visually explicit to others. The following explains the observations regarding proximity cues, highlighting the design implication that relative video size and position provide two different proximity cues that may serve different purposes for turn-taking; *relative position* can be used for addressing individuals whereas *relative size* can be used for addressing groups.

5.3.1 Resizing as a Cue for Grabbing Group Attention. Video resizing served as a way to grab attention, which extends the concept of Perceived Proxemic Distance [18, 27] from static configuration of engagement to gradual transitions between levels of engagement. Aside from configuring their video feeds to the center, a more frequent way to enhance participation in a conversation is to subtly enlarge their video feeds. A reverse pattern is that users may slightly make themselves smaller when they want to indicate that their speaking turn was finished.

5.3.2 Close Relative Distance Between Person Videos to Address Individuals. While Perceived Proxemic Distance is originally concerned with the impact of video size on interpersonal relations [18, 27], we found that *the relative position between people's video feeds* also appears to serve as a proximity cue to grab others' attention. What is especially interesting about this, is that in terms of turn-taking, it may provide a substitute for mutual gaze [36, 75], which we use as a primary cue for addressing individuals when

face-to-face. Modern meeting tools such as Gather [1] and Wonder [7] use proximity (i.e., the relative distance between videos) to determine the social group boundary and support approach and leave-taking [24]. With OpenMic, we have demonstrated the novel possibilities of using these same interaction principles but for the purpose of conveying clearer intentions between who is talking to whom within a conversation.

5.3.3 Addressing Groups vs. Individuals. The patterns of use show that video resizing is often used when grabbing the attention (lowering perceived distance) of an entire group, whereas moving one's video feed closer in relative position to another video feed means lowering the perceived distance to a specific person (which is useful for addressing a specific individual). In Gibson's terms of turn-taking [21], this then means that video resizing provides support for *turn-claiming* (i.e., addressing a group), whereas relative video positions provide support for *turn-receiving* (i.e., addressing an individual).

5.4 Limitations and Future Work

We finally turn to a discussion of the limitations of the OpenMic system and user study.

5.4.1 Scalability of OpenMic. We designed OpenMic to support turn-taking in small to medium-sized group conversations (4-10 people), and the user study only investigated the use of OpenMic using 4- and 8-person groups. While the concept of using proxemic metaphors for conversational floor transitions is not limited to smaller meetings, the system may need further design iterations to support larger group meetings. For example, the off-floor space may not be sufficient to accommodate videos of a large audience, and the floor boundary for moderated conversation may cause challenges in understanding the order of bystanders lined up for participation. A further study after design iterations with larger meetings will provide an understanding of how a video meeting interface designed using proxemic metaphors can support the transitions across groups of audience and speakers. Moreover, OpenMic mainly focuses on the proxemic metaphors of person space and digital task space. Future work can utilize and explore the effect of proxemic metaphors on turn-taking for physical task spaces such as conversations around shared physical artifacts [38].

5.4.2 Effort of Making Gradual Transitions. While our results suggest that the self-controlled gradual transitions can help users express and understand turn-taking intentions, it also showed that the interaction for gradually relocating video feeds might degrade the efficiency of transitions. This problem can be more severe if the users do not use an efficient pointing device or are not skilled with pointing interactions, as small delays in transitions can be detrimental in fluid conversational transitions [61]. To ease the transition effort, future work may investigate template layouts (such as with Gather [1] or Sprout [6]) for different kinds of turn-taking needs, or implicit interaction techniques, such as detecting implicit turn-taking cues from gestures and facial expressions.

5.4.3 On-Floor Backchannel. In OpenMic, speakers on the floor take multiple roles of making the conversation and managing the floor to control audience participation, which caused several on-floor participants to struggle to properly moderate the conversation.

In addition to the moderation burden, the confusion among the speakers about the current moderation status made it challenging for them to efficiently manage the floor. This may imply the need for a backchannel (e.g., a dedicated chat) for on-floor users to support fluid floor management.

5.4.4 Use of Deductive Analysis. We used the deductive method for qualitative analysis, which may cause overlooking some potential topics as it concentrates on the predefined themes. While we acknowledge its limitations, we consider this method to be adequate for our study as the main goal was to understand the similarities and differences between the theory (i.e., the proxemic metaphors) and the study findings. Future research may include more in-depth evaluations of OpenMic involving inductive analysis to thoroughly understand the afforded turn-taking behaviors in multiparty video meetings.

6 CONCLUSION

We have presented the concept and evaluation of OpenMic; a 2D virtual space for videoconferencing designed to support conversational floor transitions as explicit non-verbal cues for turn-taking in multiparty remote meetings. OpenMic offers a unique combination of Malleable Mirrors (changes to position and size) and a Virtual Floor (fixed-feature space), which conditions mirror manipulations by their spatial relation to it. We evaluated OpenMic in an exploratory study analyzing patterns in proxemic cues for conversational floor transitions. The study findings support the argument that with continuous control and awareness of users' video size and position, users have new means to read off others' intention to talk and identified turn-taking zones. Based on these results, we highlight the following implications for designing virtual 2D interfaces to support proxemic transitions: Support a variety in group sizes; *Trade-offs in free vs. curated positioning of videos*; *Support for different kinds of proximity cues*. These results point to exciting new directions for using proxemics theory to inform the design of videoconferencing interfaces for turn-taking.

ACKNOWLEDGMENTS

Special thanks to Josh Medrano and Celia Zeng for their support on videos and figures. This work was supported by the European Research Council (ERC) under the European Union Horizon 2020 research and innovation programme (grant agreement No 740548).

REFERENCES

- [1] 2021. Gather.Town. <https://gather.town/>. Accessed: 2021-04-15.
- [2] 2021. Lost at Sea. https://insight.typepad.co.uk/lost_at_sea.pdf. Accessed: 2021-09-09.
- [3] 2021. Mozilla Hubs. <https://hubs.mozilla.com/>. Accessed: 2021-04-15.
- [4] 2021. Remo. <https://remo.co/>. Accessed: 2021-04-15.
- [5] 2021. SpatialChat. <https://spatial.chat/>. Accessed: 2021-04-15.
- [6] 2021. Sprout. <https://sprout.place/>. Accessed: 2021-12-08.
- [7] 2021. Wonder.Me. <https://www.wonder.me>. Accessed: 2021-12-08.
- [8] 2021. Zoom. <https://zoom.us/>. Accessed: 2021-04-15.
- [9] 2022. Teamflow. <https://www.teamflowhq.com/>. Accessed: 2022-04-15.
- [10] Jeremy N Bailenson, Jim Blascovich, Andrew C Beall, and Jack M Loomis. 2003. Interpersonal Distance in Immersive Virtual Environments. *Personality and Social Psychology Bulletin* 29, 7 (2003), 819–833. <https://doi.org/10.1145/3452918.3458804>
- [11] Steve Benford, John Bowers, Lennart E Fahlén, Chris Greenhalgh, and Dave Snowden. 1995. User Embodiment in Collaborative Virtual Environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 242–249. <https://doi.org/10.1145/223904.223935>
- [12] John Bowers, James Pycoc, and Jon O'brien. 1996. Talk and Embodiment in Collaborative Virtual Environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 58–65. <https://doi.org/10.1145/238386.238404>
- [13] Bill Buxton. 2009. Mediaspace-Meaningspace-Meetingspace. In *Media Space 20+ Years of Mediated Life*. Springer, 217–231. <https://doi.org/10.1145/3170427.3173033>
- [14] William AS Buxton, Abigail J Sellen, and Michael C Sheasby. 1997. Interfaces for Multiparty Videoconferences. *Video-Mediated Communication* (1997), 385–400. <https://doi.org/10.1145/142750.143070>
- [15] Luigina Ciolli, Breda Gray, and Aparecido Fabiano Pinatti de Carvalho. 2020. Making Home Work Places. (2020).
- [16] Herbert H Clark and Susan E Brennan. 1991. Grounding in Communication. <https://doi.org/10.1109/9781118098974.ch9>
- [17] Herbert H Clark and Edward F Schaefer. 1989. Contributing to Discourse. *Cognitive Science* 13, 2 (1989), 259–294. https://doi.org/10.1207/s15516709cog130_7
- [18] Michael E Ellis. 1992. Perceived Proxemic Distance and Instructional Videoconferencing: Impact on Student Performance and Attitude. (1992).
- [19] Randi A Engle, Jennifer M Langer-Osuna, and Maxine McKinney de Royston. 2014. Toward a Model of Influence in Persuasive Discussions: Negotiating Quality, Authority, Privilege, and Access Within a Student-Led Argument. *Journal of the Learning Sciences* 23, 2 (2014), 245–268. <https://doi.org/10.1080/10508406.2014.883979>
- [20] Robert S. Fish, Robert E. Kraut, and Barbara L. Chalfonte. 1990. The VideoWindow System in Informal Communication. In *Proceedings of the 1990 ACM Conference on Computer-Supported Cooperative Work* (Los Angeles, California, USA) (CSCW '90). Association for Computing Machinery, New York, NY, USA, 1–11. <https://doi.org/10.1145/99332.99335>
- [21] David R Gibson. 2003. Participation Shifts: Order and Differentiation in Group Conversation. *Social Forces* 81, 4 (2003), 1335–1380. <https://doi.org/10.1353/sof.2003.0055>
- [22] David R Gibson. 2005. Taking Turns and Talking Ties: Networks and Conversational Interaction. *Amer. J. Sociology* 110, 6 (2005), 1561–1597. <https://doi.org/10.1086/428689>
- [23] Erving Goffman. 1981. *Forms of Talk*. University of Pennsylvania Press.
- [24] Carlos Gonzalez Diaz, John Tang, Advait Sarkar, and Sean Rintel. 2022. Making Space for Social Time: Supporting Conversational Transitions Before, During, and After Video Meetings. In *2022 Symposium on Human-Computer Interaction for Work* (Durham, NH, USA) (CHIWORK 2022). Association for Computing Machinery, New York, NY, USA, Article 4, 11 pages. <https://doi.org/10.1145/3533406.3533417>
- [25] Mar Gonzalez-Franco, Mark Hall, Devon Hansen, Karl Jones, Paul Hannah, and Pablo Bermell-Garcia. 2015. Framework for Remote Collaborative Interaction in Virtual Environments Based On Proximity. In *2015 IEEE Symposium on 3D User Interfaces (3DUI)*. 153–154. <https://doi.org/10.1109/3DUI.2015.7131746>
- [26] Jack Arthur Gowan and James Michael Downs. 1994. Video Conferencing Human-Machine Interface: a Field Study. *Information & Management* 27, 6 (1994), 341–356. [https://doi.org/10.1016/0378-7206\(94\)90015-9](https://doi.org/10.1016/0378-7206(94)90015-9)
- [27] David Michael Grayson. 2000. *The Role of Perceived Proximity in Video-Mediated Communication*. Ph.D. Dissertation. University of Glasgow. <https://doi.org/10.1145/565711.565713>
- [28] Saul Greenberg, Nicolai Marquardt, Till Ballendat, Rob Diaz-Marino, and Miaosen Wang. 2011. Proxemic Interactions: the New Ubicomp? *Interactions* 18, 1 (2011), 42–50. <https://doi.org/10.1145/1897239.1897250>
- [29] Jens Emil Grønbaek, Mille Skovhus Knudsen, Kenton O'Hara, Peter Gall Krogh, Jo Vermeulen, and Marianne Graves Petersen. 2020. *Proxemics Beyond Proximity: Designing for Flexible Social Interaction Through Cross-Device Interaction*. Association for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/3313831.3376379>
- [30] Jens Emil Grønbaek, Henrik Korsgaard, Marianne Graves Petersen, Morten Henriksen Birk, and Peter Gall Krogh. 2017. Proxemic Transitions: Designing Shape-Changing Furniture for Informal Meetings. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) (CHI '17). Association for Computing Machinery, New York, NY, USA, 7029–7041. <https://doi.org/10.1145/3025453.3025487>
- [31] Jens Emil Grønbaek, Banu Saatçi, Carla F. Griggio, and Clemens Nylandstedt Klokmose. 2021. MirrorBlender: Supporting Hybrid Meetings With a Malleable Video-Conferencing System. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 451, 13 pages. <https://doi.org/10.1145/3411764.3445698>
- [32] Sofiane Gueddana and Nicolas Roussel. 2006. PêLe-MêLe, a Video Communication System Supporting a Variable Degree of Engagement. In *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work* (Banff, Alberta, Canada) (CSCW '06). Association for Computing Machinery, New York, NY, USA, 423–426. <https://doi.org/10.1145/1180875.1180938>
- [33] Sofiane Gueddana and Nicolas Roussel. 2006. PêLe-MêLe, a Video Communication System Supporting a Variable Degree of Engagement. In *Proceedings of the 2006*

- 20th Anniversary Conference on Computer Supported Cooperative Work (Banff, Alberta, Canada) (CSCW '06). Association for Computing Machinery, New York, NY, USA, 423–426. <https://doi.org/10.1145/1180875.1180938>
- [34] Carl Gutwin. 2002. Traces: Visualizing the Immediate Past to Support Group Interaction. In *Graphics Interface*. Citeseer, Citeseer, 43–50. <https://doi.org/10.1145/3105971.3105981>
- [35] Edward Twitchell Hall. 1966. *The Hidden Dimension*. Vol. 609. Garden City, NY: Doubleday.
- [36] Zhenyi He, Keru Wang, Brandon Yushan Feng, Ruofei Du, and Ken Perlin. 2021. GazeChat: Enhancing Virtual Conferences With Gaze-Aware 3D Photos. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. 769–782. <https://doi.org/10.1145/3472749.3474785>
- [37] Erzhen Hu, Md Asshikur Rahman Azim, and Seongkook Heo. 2022. FluidMeet: Enabling Frictionless Transitions Between In-Group, Between-Group, and Private Conversations During Virtual Breakout Meetings. In *Proceedings of the 40th Annual ACM Conference on Human Factors in Computing Systems* (New Orleans, USA) (CHI '22). Association for Computing Machinery, New York, NY, USA, 17 pages. <https://doi.org/10.1145/3491102.3517558>
- [38] Erzhen Hu, Jens Emil Grønbaek, Wen Ying, Ruofei Du, and Seongkook Heo. 2023. ThingShare: Ad-Hoc Digital Copies of Physical Objects for Sharing Things in Video Meetings. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA. <https://doi.org/10.1145/3544548.3581148>
- [39] Hiroshi Ishii and Minoru Kobayashi. 1992. Clearboard: a Seamless Medium for Shared Drawing and Conversation With Eye Contact. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 525–532. <https://doi.org/10.1145/142750.142977>
- [40] Shahram Izadi, Ankur Agarwal, Antonio Criminisi, John Winn, Andrew Blake, and Andrew Fitzgibbon. 2007. C-Slate: Exploring Remote Collaboration on Horizontal Multi-Touch Surfaces. (2007).
- [41] Tracy Jenkin, Jesse McGeachie, David Fono, and Roel Vertegaal. 2005. EyeView: Focus+ Context Views for Large Group Video Conferences. In *CHI'05 Extended Abstracts on Human Factors in Computing Systems*. 1497–1500. <https://doi.org/10.1145/1056808.1056950>
- [42] Adam Kendon. 1990. *Conducting Interaction: Patterns of Behavior in Focused Encounters*. Vol. 7. CUP Archive.
- [43] Peter Gall Krogh, Marianne Graves Petersen, Kenton O'Hara, and Jens Emil Grønbaek. 2017. Sensitizing Concepts for Socio-Spatial Literacy in HCI. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) (CHI '17). Association for Computing Machinery, New York, NY, USA, 6449–6460. <https://doi.org/10.1145/3025453.3025756>
- [44] Celine Latulipe. 2021. *A CS1 Team-Based Learning Space in Gather.Town*. Association for Computing Machinery, New York, NY, USA, 1245. <https://doi.org/10.1145/3408877.3439587>
- [45] Bryan Lawson. 2007. *Language of Space*. Routledge.
- [46] Bokyung Lee, Michael Lee, Pan Zhang, Alexander Tessier, and Azam Khan. 2019. An empirical study of how socio-spatial formations are influenced by interior elements and displays in an office context. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–26.
- [47] Paul Luff, Christian Heath, Hideaki Kuzuoka, Keiichi Yamazaki, and Jun Yamashita. 2006. Handling Documents and Discriminating Objects in Hybrid Spaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 561–570. <https://doi.org/10.1145/1124772.1124858>
- [48] Nicolai Marquardt, Till Ballendat, Sebastian Boring, Saul Greenberg, and Ken Hinckley. 2012. Gradual Engagement: Facilitating Information Exchange Between Digital Devices As a Function of Proximity. In *Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces*. 31–40. <https://doi.org/10.1145/2396636.2396642>
- [49] Nicolai Marquardt, Robert Diaz-Marino, Sebastian Boring, and Saul Greenberg. 2011. The Proximity Toolkit: Prototyping Proxemic Interactions in Ubiquitous Computing Ecologies. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology* (Santa Barbara, California, USA) (UIST '11). Association for Computing Machinery, New York, NY, USA, 315–326. <https://doi.org/10.1145/2047196.2047238>
- [50] Nicolai Marquardt, Nathalie Henry Riche, Christian Holz, Hugo Romat, Michel Pahud, Frederik Brudy, David Ledo, Chunjong Park, Molly Jane Nicholas, Teddy Seyed, et al. 2021. AirConstellations: In-Air Device Formations for Cross-Device Interaction Via Multiple Spatially-Aware Armatures. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. 1252–1268. <https://doi.org/10.1145/3472749.3474820>
- [51] Nicolai Marquardt, Ken Hinckley, and Saul Greenberg. 2012. Cross-Device Interaction Via Micro-Mobility and F-Formations. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology*. 13–22. <https://doi.org/10.1145/2380116.2380121>
- [52] Osamu Morikawa and Takanori Maesaka. 1998. HyperMirror: Toward Pleasant-to-Use Video Mediated Communication System. In *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work*. 149–158. <https://doi.org/10.1145/289444.289489>
- [53] Archana Narayanan, Erzhen Hu, and Seongkook Heo. 2022. Enabling Remote Hand Guidance in Video Calls Using Directional Force Illusion. In *Companion Publication of the 2022 Conference on Computer Supported Cooperative Work and Social Computing* (Virtual Event, Taiwan) (CSCW'22 Companion). Association for Computing Machinery, New York, NY, USA, 135–139. <https://doi.org/10.1145/3500868.3559470>
- [54] Caleece Nash, Mohammad Hossein Jarrahi, and Will Sutherland. 2021. Nomadic Work and Location Independence: the Role of Space in Shaping the Work of Digital Nomads. *Human Behavior and Emerging Technologies* 3, 2 (2021), 271–282.
- [55] David Nguyen and John Canny. 2005. MultiView: Spatially Faithful Group Video Conferencing. Association for Computing Machinery, New York, NY, USA, 799–808. <https://doi.org/10.1145/1240624.1240846>
- [56] Brid O'Conaill, Steve Whittaker, and Sylvia Wilbur. 1993. Conversations Over Video Conferences: an Evaluation of the Spoken Aspects of Video-Mediated Communication. *Human-Computer Interaction* 8, 4 (1993), 389–428. https://doi.org/10.1207/s15327051hci080_4
- [57] Kenton O'hara, Jesper Kjeldskov, and Jeni Paay. 2011. Blended Interaction Spaces for Distributed Team Collaboration. *ACM Trans. Comput.-Hum. Interact* 18, 1, Article 3 (2011), 28 pages. <https://doi.org/10.1145/1959022.1959025>
- [58] Riëks op den Akker, Dennis Hof, Hendri Hondorp, Harm op den Akker, Job Zwiers, and Anton Nijholt. 2009. Supporting Engagement and Floor Control in Hybrid Meetings. In *Cross-Modal Analysis of Speech, Gestures, Gaze and Facial Expressions*. Springer, 276–290. https://doi.org/10.1007/978-3-642-03320-9_26
- [59] Alexandre Pauchet, François Coldefy, Liv Lefebvre, S Louis Ditt Picard, Arnaud Bouguet, Laurence Perron, Joël Guerin, Daniel Corvaisier, and Michel Collobert. 2007. Mutual Awareness in Collocated and Distant Collaborative Tasks Using Shared Interfaces. In *IFIP Conference on Human-Computer Interaction*. Springer, 59–73.
- [60] Mark Perry, Kenton O'hara, Abigail Sellen, Barry Brown, and Richard Harper. 2001. Dealing With Mobility: Understanding Access Anytime, Anywhere. *ACM Transactions on Computer-Human Interaction (TOCHI)* 8, 4 (2001), 323–347. <https://doi.org/10.1145/504704.504707>
- [61] Sean Rintel, Abi Sellen, Advait Sarkar, Priscilla Wong, Nancy Baym, and Rachel Bergmann. 2020. Study of Microsoft Employee Experiences in Remote Meetings During COVID-19 (Project Tahiti). Microsoft Research. Retrieved from [https://www.microsoft.com/en-us ...](https://www.microsoft.com/en-us...), Microsoft Research. Retrieved from [https://www.microsoft.com/en-us ...](https://www.microsoft.com/en-us...)
- [62] Roy Rodenstein and Judith S. Donath. 2000. Talking in Circles: Designing a Spatially-Grounded Audioconferencing Environment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (The Hague, The Netherlands) (CHI '00). Association for Computing Machinery, New York, NY, USA, 81–88. <https://doi.org/10.1145/332040.332410>
- [63] Nicolas Roussel, Helen Evans, and Heiko Hansen. 2004. MirrorSpace: Using Proximity as an Interface to Video-Mediated Communication. 345–350. https://doi.org/10.1007/978-3-540-24646-6_25
- [64] Stacey D Scott, M Sheelagh T Carpendale, and Kori Inkpen. 2004. Territoriality in Collaborative Tabletop Workspaces. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work*. 294–303. <https://doi.org/10.1145/1031607.1031655>
- [65] Abigail Sellen, Bill Buxton, and John Arnott. 1992. Using Spatial Cues to Improve Videoconferencing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Monterey, California, USA) (CHI '92). Association for Computing Machinery, New York, NY, USA, 651–652. <https://doi.org/10.1145/142750.143070>
- [66] Abigail J Sellen. 1995. Remote Conversations: the Effects of Mediating Talk With Technology. *Human-Computer Interaction* 10, 4 (1995), 401–444. https://doi.org/10.1207/s15327051hci100_2
- [67] Mauricio Sousa, Daniel Mendes, Daniel Medeiros, Alfredo Ferreira, João Madeiras Pereira, and Joaquim Jorge. 2016. Remote Proxemics. In *Collaboration Meets Interactive Spaces*. Springer, 47–73. https://doi.org/10.1007/978-3-319-45853-4_4
- [68] Mark Stefik, Daniel G Bobrow, Gregg Foster, Stan Lanning, and Deborah Tatar. 1987. WYSIWIS Revised: Early Experiences With Multiuser Interfaces. *ACM Transactions on Information Systems (TOIS)* 5, 2 (1987), 147–167. <https://doi.org/10.1145/27636.28056>
- [69] Anthony Tang, Michel Pahud, Kori Inkpen, Hrvoje Benko, John C Tang, and Bill Buxton. 2010. Three's Company: Understanding Communication Channels in Three-Way Distributed Collaboration. In *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work*. 271–280. <https://doi.org/10.1145/1718918.1718969>
- [70] John C Tang. 1991. Findings From Observational Studies of Collaborative Work. *International Journal of Man-Machine Studies* 34, 2 (1991), 143–160. <https://doi.org/10.1016/0020>
- [71] John C Tang, Ellen A Isaacs, and Monica Rua. 1994. Supporting distributed groups with a montage of lightweight interactions. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work*. 23–34. <https://doi.org/10.1145/192844.192861>
- [72] John C. Tang and Monica Rua. 1994. Montage: Providing Teleproximity for Distributed Groups. In *Proceedings of the SIGCHI Conference on Human Factors*

- in *Computing Systems* (Boston, Massachusetts, USA) (*CHI '94*). Association for Computing Machinery, New York, NY, USA, 37–43. <https://doi.org/10.1145/191666.191688>
- [73] Maria Tomprou, Young Ji Kim, Prerna Chikersal, Anita Williams Woolley, and Laura A Dabbish. 2021. Speaking Out of Turn: How Video Conferencing Reduces Vocal Synchrony and Collective Intelligence. *PloS One* 16, 3 (2021), e0247655.
- [74] Philip Tuddenham and Peter Robinson. 2009. Territorial Coordination and Workspace Awareness in Remote Tabletop Collaboration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) (*CHI '09*). Association for Computing Machinery, New York, NY, USA, 2139–2148. <https://doi.org/10.1145/1518701.1519026>
- [75] Roel Vertegaal. 1999. The GAZE Groupware System: Mediating Joint Attention in Multiparty Communication and Collaboration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 294–301. <https://doi.org/10.1145/302979.303065>
- [76] Roel Vertegaal, Ivo Weevers, and Changuk Sohn. 2002. GAZE-2: an Attentive Video Conferencing System. In *CHI'02 Extended Abstracts on Human Factors in Computing Systems*. 736–737. <https://doi.org/10.1145/506443.506572>
- [77] John M Wiemann and Mark L Knapp. 2017. Turn-Taking in Conversations. *Communication Theory* (2017), 226–245. <https://doi.org/10.1109/ROMAN.2014.6926266>
- [78] Julie Williamson, Jie Li, Vinoba Vinayagamoorthy, David A. Shamma, and Pablo Cesar. 2021. Proxemics and Social Interactions in an Instrumented Virtual Reality Workshop. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (*CHI '21*). Association for Computing Machinery, New York, NY, USA, Article 253, 13 pages. <https://doi.org/10.1145/3411764.3445729>
- [79] Bin Xu, Jason Ellis, and Thomas Erickson. 2017. Attention From Afar: Simulating the Gazes of Remote Participants in Hybrid Meetings. In *Proceedings of the 2017 Conference on Designing Interactive Systems*. 101–113. <https://doi.org/10.1145/3064663.3064720>
- [80] Xiaolong Zhang and George W. Furnas. 2002. Social Interactions in Multiscale CVEs. In *Proceedings of the 4th International Conference on Collaborative Virtual Environments* (Bonn, Germany) (*CVE '02*). Association for Computing Machinery, New York, NY, USA, 31–38. <https://doi.org/10.1145/571878.571884>