HyWay: Enabling Mingling in the Hybrid World*

HARSH VIJAY†‡#, Microsoft Research, India
SAUMAY PUSHP‡#§, Microsoft Research, India
AMISH MITTAL‡#, Microsoft Research, India
PRAVEEN GUPTA†#, University of British Columbia, Canada
MEGHNA GUPTA¶, Microsoft Research, India
SIRISH GAMBHIRA‡#, Microsoft Research, India
SHIVANG CHOPRA‡#, Georgia Institute of Technology, USA
MAYANK BARANWAL†#, Carnegie Mellon University, USA
ARSHIA ARYA¶, Microsoft Research, India
AJAY MANCHEPALLI**, Microsoft Research, India
VENKATA N. PADMANABHAN+, Microsoft Research, India

We present HyWay, short for "**Hy**brid Hall**way**", to enable mingling and informal interactions among physical and virtual users, in casual spaces and settings, such as office water cooler areas, conference hallways, trade show floors, and more. We call out how the *hybrid* and *unstructured (or semi-structured)* nature of such settings set these apart from the all-virtual and/or structured settings considered in prior work. Key to the design of HyWay is bridging the *awareness gap* between physical and virtual users, and providing the virtual users the same *agency* as physical users.

To this end, we have designed HyWay to incorporate *reciprocity* (users can see and hear others only if they can be seen and heard), *porosity* (conversations in physical space are porous and not within airtight compartments), and *agency* (the ability for users to seamlessly move between conversations). We present our implementation of HyWay and the user survey findings from multiple deployments in unstructured settings (e.g., social gatherings), and semi-structured ones (e.g., a poster event). Results from these deployments show that HyWay enables effective mingling between physical and virtual users.

CCS Concepts: • **Human-centered computing → Collaborative and social computing systems and tools**;

*All the authors were with Microsoft Research India during their involvement in this work. The authors, other than the last two, are listed in reverse alphabetical order. They contributed in various ways, including:
†: building the initial, functional HyWay system.
‡: building subsequent refinements and augmentations to the HyWay system.
#: deploying HyWay in various settings.
§: conducting user studies.
¶: creating maps and visualization.
As for the last two authors, their roles were:
**: contributing to the design of HyWay and being the primary driver of the deployments.
+: conceiving HyWay and its main design elements, and leading the overall project.

Authors' email addresses: Harsh Vijay, harsh.vijay@microsoft.com; Saumay Pushp, saumay.pushp@microsoft.com; Amish Mittal, t-ammittal@microsoft.com; Praveen Gupta, pvgupta@student.ubc.ca; Meghna Gupta, t-meggupta@microsoft.com; Sirish Gambhira, t-sigambhira@microsoft.com; Shivang Chopra, shivangchopra11@gatech.edu; Mayank Baranwal, mbaranwa@andrew.cmu.edu; Arshia Arya, t-aarya@microsoft.com; Ajay Manchepalli, ajayma@microsoft.com; Venkata N. Padmanabhan, padmanab@microsoft.com.

2474-9567/2023/6-ART77 $15.00
https://doi.org/10.1145/3596235

Additional Key Words and Phrases: Hybrid mingling, unstructured and semi-structured conversations, awareness, agency, porosity, reciprocity

ACM Reference Format:
Harsh Vijay, Saumay Pushp, Amish Mittal, Praveen Gupta, Meghna Gupta, Sirish Gambhira, Shivang Chopra, Mayank Baranwal, Arshia Arya, Ajay Manchepalli, and Venkata N. Padmanabhan. 2023. HyWay: Enabling Mingling in the Hybrid World. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 2, Article 77 (June 2023), 33 pages. https://doi.org/10.1145/3596235

1 INTRODUCTION

Remote work became the norm during the extended work from home (WFH) period arising from the COVID-19 pandemic. Tools such as Teams [25] and Zoom [30] became the lifeline of work meetings, which tend to be structured — scheduled, with an agenda, a predetermined set of invitees, and often with just a few dominant speakers (or even one, as in a lecture). The need to support unstructured interactions, like hallway conversations at a conference or chit-chat by the office water cooler, led to the development of newer "spatial" tools such as Gather [13] and SpatialChat [23], but these are confined to users in virtual-only settings.

With the pandemic waning, employees have been returning to the physical workplace. However, the convenience and flexibility of WFH have not been forgotten. This combination has led to the emergence of the hybrid workplace, with a mix of in-person[1] and remote[2] users. In some cases, a subset of employees WFH, and in other cases many or all employees WFH on specific days of the week. Hybrid has become the norm for conferences and other events too, with remote attendance providing users the opportunity to participate without the hassle or expense (or incurring the carbon footprint) of travel.

The hybrid workplace (or event) setting, however, has brought to the fore a key challenge that is exemplified by the following quote from an article titled "Staying Visible When Your Team Is in the Office…But You're WFH" [24]: *"...if you plan to work remotely full-time or most of the time, how can you stay visible when your in-office colleagues are likely to have far more exposure ... Colleagues working together in an office have plenty of organic opportunities, from elevator rides to breakroom encounters ..."*. Existing tools fall short either in not supporting *unstructured* interactions, or in not supporting *hybrid* settings, or both.

In this paper, we present the design and implementation of and the deployment experience with HyWay, short for "**Hy**brid Hall**way**", which is a system to support unstructured and semi-structured, hybrid interactions. The goal of HyWay is to enable the interactions and mingling that happen in a physical setting such as hallway, but in a situation where some of the users are remote.

Users tend to be comfortable with physical interactions and mingling in unstructured settings because these tend to be surprise-free and fluid. *Awareness* ensures that users know of the presence of others around, not just those they are talking to but also others in the vicinity. They are often also able to overhear nearby conversation(s). If someone they see or something they overhear catches their interest, users have the *agency* to move over and join that conversation.

The key question is how best can we enable remote users to "plug in" while retaining the surprise-free and flexible nature of such interactions. In HyWay, remote users join using large displays that are placed at various locations in the hallway [3] (which we term as "physical zones"), with the ability to move seamlessly between the zones using a map-based interface. Very little changes for in-person users — they engage with each other just as in a purely-physical setting except that when they are in the vicinity of a display, they would *additionally* have the opportunity to engage with remote users. Figure 1 shows an example HyWay deployment with 3 screens and the respective HyWay zones.

[1] "in-person" and "physical" user(s) are used interchangeably throughout the paper.
[2] "remote" and "virtual" user(s) are used interchangeably throughout the paper.
[3] For ease of exposition and in keeping with our project name, we use the term "hallway" to generically refer to all informal gathering spaces, including the lobby, atrium, cafeteria, etc.

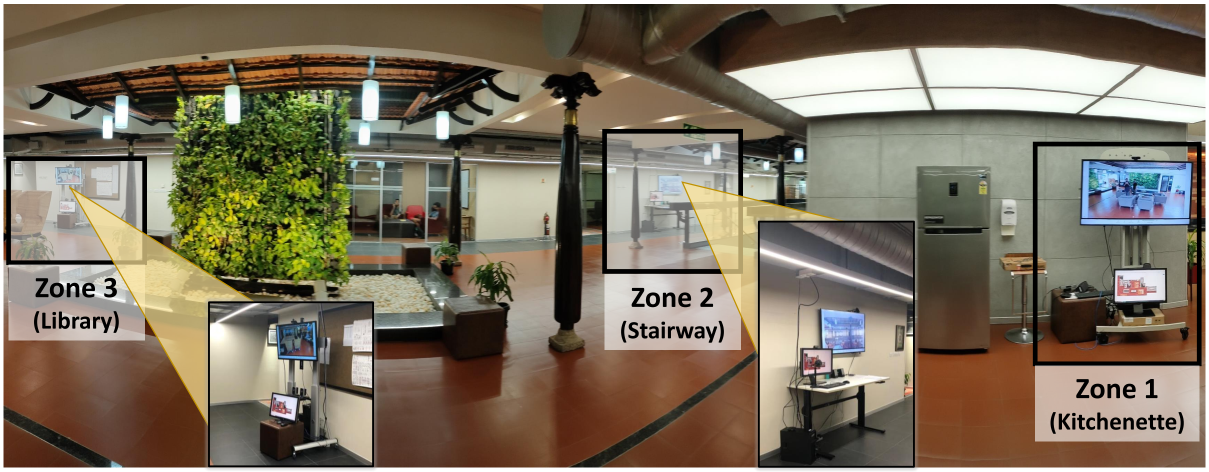


Fig. 1. HyWay deployed in an office workspace. Each setup comprises of a 55 inch display (used to interact with remote users), and a 24 inch display (shows the presence and movement of remote users on the map), wireless mic (Sennheiser SK-100 G3 [22]), speakers (Creative SBS-240 [6]) and powered by a 3.4 GHz processor with 32 GB RAM.

To bridge the *awareness gap* between the in-person and remote users, HyWay employs three design elements: reciprocity, porosity, and map-based awareness. *Reciprocity* ensures that users — both in-person and remote — can see and hear others only if they are themselves seen and heard. *Porosity* ensures that users engaged in a conversation can overhear nearby conversations. *Map-based awareness* uses a skeuomorphic representation of space, allowing users to be aware of others nearby.

Together, the above ensures that users are *not* locked into airtight conversations. Instead, users (both in-person and remote) are fully aware of the presence of other nearby in-person and remote users, both through aural (porosity) and visual (map) cues. This naturally facilitates *agency*. That is, just as in-person users in the hallway are at liberty to move about, so can remote users move from one conversation to another using the map-based interface.

We make the following contributions in this paper:

(1) The design of HyWay and its novel elements compared to prior systems, including reciprocity and porosity.
(2) The implementation of HyWay as a browser-based app.
(3) The learnings from multiple HyWay deployments and corresponding user surveys.

2 RELATED WORK

Studies have shown that ad-hoc, informal conversations are key to creative innovation in social and professional contexts and are a critical factor when forming social bonds and building trust in a group [45, 49]. Here, we present a survey of prior work on enabling such conversations.

Background on unstructured interactions in physical workplaces. Kraut, et. al [56] distinguished four categories of interactions: planned (prearranged meetings), intended (explicitly sought by one person), opportunistic (anticipated by one party but occurring only when the parties happen to see each other), and spontaneous (unanticipated by either party). We focus on these latter two types of interactions, which we together term as "unintended." [56] estimated that unintended conversations made up 52% of the interactions that occurred in the workplace they studied and accounted for much of the information flow there.

In another study, Whittaker, et al. [74] found that 92% of the interactions they observed in two office settings were not pre-arranged. These interactions were frequent and very short, with each interaction lasting under two minutes on average. Furthermore, past study [57] have shown that the more a group engages in unplanned interactions, the more productive they are in their respective projects. In fact, a lack of unstructured interactions reduces the level of coordination and progress on projects [56].

Systems supporting unstructured interactions. Despite the critical role of unstructured interactions, only a few systems have included features that make it possible to come across someone unintentionally and start up a conversation. The media space work ([36], [42]) highlighted the use of audio and video connections among distributed sites to help people notice the work activity of others in remote sites and make spontaneous contact. However, explicit attempts to support serendipitous encounters through technology has so far been quite limited. The main shortcoming of past attempts is that they have tried to bring together people who do not necessarily have the context or the need to interact [56].

One media space system that did gain acceptance is Portholes at Xerox EuroPARC [39]. It allowed people to stay aware of others by viewing a matrix of slowly updating video snapshots of the offices of a fixed set of people. These views could prompt a user to establish a direct audio-video call with someone they saw in Portholes.

The hybrid workplace. Telecommuting is an old idea that dates back to at least the 1970s [69]. However, the shift from physical to digital accelerated during the COVID-19 pandemic in 2020-2021, leading to organizations adopting hybrid or even entirely digital ways to do their work [72]. Many companies believe that the post-pandemic future of work will be hybrid. For example, Microsoft's 2021 Work Trend Index Annual Report finds that 66% of leaders say that their company is considering redesigning office space for hybrid work and 73% of remote employees want flexible remote work options to continue [26].

Prior work on tele-presence for meetings has explored many variations in both social and technological factors ([59],[42]). The main focal point of our work arises from the hybrid collaboration defined by Neumayr et al. [64]. Hybrid ecosystems often comprise an interwoven mix of space, devices, and software, both for a single individual (e.g., their home and office set-up) and for teams working together. One example of the initial steps in providing specific support for such hybrid spaces is MirrorBlender [47], which provides a malleable videoconferencing system for hybrid meetings. The authors suggest a WYSIWIS (What-You-See-Is-What-I-See) style 2D frame such that every participant can see the same layout of mirrors, i.e., the video feeds, from other remote participants. The goal of the study was to enable synchronizing and blending of different camera perspectives together such that it would become easy for the participants to point at specific elements on a shared screen using their own camera image for "pointing gestures".

In this paper, we investigate social interactions beyond work-related tasks among workers in a hybrid work setting. For example, we consider small conversations that happen at the coffee machines. We envision such conversations happening between in-person users and remote users. Opportunistic (spontaneous and serendipitous) and informal interactions throughout the workday play a major role in coordination, productivity, and the well-being of groups [61]. Opportunistic talk happens at lunches [33], in hallways [60], and by water-coolers [58]. This makes users gossip [65], have over-the-shoulder-learning [73], and have productive side conversations even while seated at their workplace desks [34].

Current technologies and interfaces for unstructured meetings. It is instructive to place HyWay in the context of existing systems. Unlike systems such as Teams [25] and Zoom [30], which do not provide users any visibility into the goings-on beyond the call they are in, HyWay provides broad awareness. While systems such as SpatialChat [23] and Gather [13] provide awareness and navigation using a map interface, an approach we also use in HyWay, there are crucial differences. Unlike the point-and-click interface of SpatialChat and the ability for users to teleport themselves into a distant conversation, HyWay ensures that users can only approach gradually, just as they would in the physical setting, which helps ensure a surprise-free interaction. Furthermore, unlike both

SpatialChat and Gather, HyWay does not compartmentalize the space into discrete rooms but instead provides porosity (with suitable attenuation, as discussed in Section 5) to allow users to overhear multiple conversations, just as it would happen in a physical setting. Finally, unlike these other systems, which are for a purely virtual setting, HyWay operates in a hybrid setting, so the map used by remote users to navigate around corresponds to the space that the physical users are in and so, for instance, even the overhearing of conversations is not arbitrary but rather mimics what would have happened if a remote user was present physically on the floor.

Most videoconferencing platforms, such as Zoom [30], provide breakout rooms for enabling unstructured mingling among participants. Users can have a focused in-group conversations and join other breakout rooms with an awareness of the other rooms' titles and membership. A few other platforms such as Discord [7], Unhangout [28], and Remo [21] have provided advanced visualizations of the other rooms' activities (e.g., icons next to user avatars or usernames) to allow out-of-group members (i.e., users who are not part of a conversation) to identify the active speaker in a room or which other out-of-group members are in the room. Encouraged by these approaches that provide additional context on out-of-group activities in the vicinity, our work focuses on building awareness among both physical and virtual users by providing a skeuomorphic map-based interface, and the ability to overhear nearby conversations, thereby encouraging users to join such conversations.

Providing awareness in unstructured computer-supported virtual meetings. The challenges of feeling disconnected and losing conversation context occurs not only in unstructured meetings, but also during meetings that are held using computer-mediated communication platforms [36] that connect distributed teams and facilitate informal social interactions. Consequently, this lack of awareness, as noted in [39], would lead to missed opportunities to either collaborate or mingle with others [43].

To understand awareness in workplaces, previous researchers have primarily focused on studying asynchronous communication tools such as email or instant messaging [48]. Several other studies have also focused on understanding the benefits of peeking into conversations, e.g., in the context of text-based instant messaging ([41], [40]) and chat rooms. To harness the power of casual conversations during workspace collaborations [55], group-aware systems have created small social worlds or communities through large displays or personal workstations ([62], [46]). These systems have the potential to enable new collaborations across teams by supporting shared video snapshots of offices (e.g., [39]) and lightweight peeks into other offices with a sense of tele-proximity (e.g., [50]). These works enabled distant teams to transitions from weak to strong coupling by providing casual communication opportunities.

Tele-presence robots in a hybrid world. The ability to move around a physical space through a tele-presence robot is powerful. Such robots provide a sense of presence and agency to remote users as they can move around more actively and freely in a physical space (e.g., HomeMeld [51]). Past research has explored the engineering and human factors challenges to realizing the use of tele-presence robots in various contexts ([37], [66], [53], [67]) involving collaborative [68] and non-collaborative [44] workplaces, respectively. For example, robotic tele-presence has found use in settings such as classroom learning [70].

Tele-presence robots and HyWay present an interesting trade-off and hence the opportunity to leverage the complementary strengths of each. Robots, especially human-sized ones [8], or better still humanoid ones [1], would help remote users make their presence much more salient to other (in-person) users than as a video tile on a display as in HyWay. Robots would also allow remote users to navigate to where they wish instead of being tied down to where the displays are located or necessitating the expense of plastering a large space with several displays.

However, these advantages of tele-presence robots come at a cost. Each robot typically costs thousands of dollars [8], which is much higher than the hundreds of dollars it would cost for a large-screen display with peripherals. Furthermore, having remote users just appear as video tiles on large displays avoids physical constraints to navigation (e.g., a wheeled robot might not be able to navigate level changes), and accommodates a

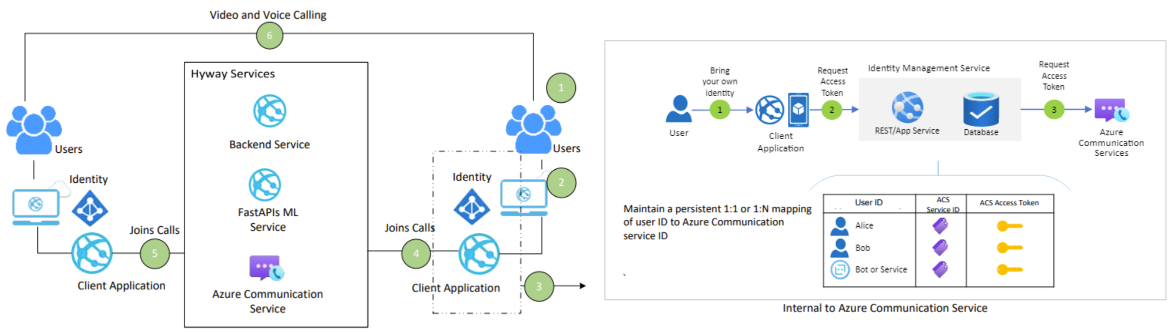


Fig. 2. Workflow between server and client in HyWay.

higher density of users (e.g., ∼ 9 human-size faces could be accommodated on a 50-inch display versus just 1 remote user per mobile robot).

We believe that the trade-off between these approaches centers on the user density. If a large area is to be covered with a relatively small number of remote users participating, it might be advantageous to have robots as the means for the remote users to navigate the space. On the other hand, if the space is compact and so the remote user density is high (as has been the case in our deployments of HyWay), it might be more cost-effective to use large, fixed displays. In general, one could employ a combination of the two approaches, e.g., having highly mobile users ride on robots while anchoring users on the fixed displays when they are stationary, say engaged in a conversation.

3 DESIGN OF HYWAY

The physical experience defines the starting point for our design of HyWay. As noted in Section 1, in a physical hallway setting, users enjoy surprise-free and fluid interactions. Our goal in HyWay is to preserve these benefits for in-person users while enabling remote users to enjoy the same benefits as best as possible.

3.1 Hardware Setup

In HyWay, we deploy large screens (along with cameras, speakers, and microphones) at set locations in the physical space (Figure 1) to enable remote users to be "present". While this choice limits the set of locations where remote users can be present, this has not proved to be a bottleneck in our deployments so far (Section 8) and offers multiple benefits.

Compared to the approach of having remote users call in to the mobile devices such as smartphones or tablets of in-person users, HyWay imposes minimally on the in-person users (who would only have to be in the vicinity of a screen if they wish to engage with the remote users) and enables a shared experience in a manner that is difficult with a small-screen personal device.

3.2 HyWay Client and Server Architecture.

HyWay, facilitates its users to join calls, on-demand, and that too without any per-call invitation link. Remote users can simply launch HyWay whenever they want using an instance-specific HyWay URL.

Under the covers, every user in HyWay has a unique user ID, which is application-specific and remains the same across multiple instances of and multiple sessions of the HyWay application. This ID is obtained from an authentication service, which is Azure Active Directory (AAD) in our implementation. Every HyWay zone, whether physical or virtual, starts up a call with a unique call ID. We use Azure Communication Services (ACS) for this purpose. When a user's client connects to a call, the communications service creates a communication service ID for that user's session.

Figure 2 illustrates the workflow of the system. (1) A unique call ID is assigned to each physical and virtual zone. However, no calls are started at this stage. (2) When a remote user joins HyWay, the user's client joins each of the per-zone calls identified by the unique call IDs (calls are initialized by ACS behind the scenes). In instances where there are many zones, for reasons of efficiency, the client joins only a subset of the calls in its vicinity (Section 6.2). (3) Now, the communication service creates and maintains a mapping between the user ID and the communication service ID, which persists during the user's HyWay session. (4)-(5) When users want to communicate with one another, they join the corresponding calls (identified by the same call IDs) to receive the audio and video feeds (6).

3.3 Bridging the Awareness Gap

As touched upon in Section 1, awareness of the presence of those around is key to enabling a surprise-free environment, where users are comfortable mingling and engaging in informal conversation. Awareness arises from a combination of visual, aural, and spatial information and cues, so the challenge in HyWay is to enable these as effectively as possible for both in-person and remote users.

Ensuring mutual awareness among in-person users does not require any special effort since this is routine in the physical hallway conversation setting. Ensuring such awareness among remote users is also not challenging given the experience with audio/video conferencing tools, although the common practice of users keeping their microphones and cameras turned off might not jibe with informal interaction. The more challenging task is ensuring awareness across in-person and remote users (bridging the "awareness gap") given the constraints of the typically small number of HyWay displays that would be deployed in the physical space.

To bridge the awareness gap, HyWay includes the following elements.

First, *reciprocity* (Section 4) makes in-person and remote users aware of each other's presence by ensuring that users can participate by seeing and hearing others only if they are themselves seen and heard by others. While remote users can often "participate" in agenda-driven online meetings (e.g., a business discussion or a university lecture) while leaving their microphone muted and camera turned off, we believe that the ability to see and hear others is key to enabling mingling in informal and typically agenda-less settings. Therefore, HyWay incorporates a combination of technical enforcement and social constraints to ensure reciprocity.

Second, *porosity* (Section 5) ensures that conversations in HyWay are not airtight compartments as online meetings typically are. Just as users in a physical hallway would be aware of and often also able to overhear bits of conversations in the vicinity, porosity enables HyWay users — both in-person and remote — to hear and be heard in multiple nearby conversations. Of course, such overhearing would need to be suitably muffled to reflect the attenuation over distance that would happen in a physical setting, where the conversation being overheard takes place some distance away from the user.

Third, HyWay uses a *map-based* visualization (Section 6) to make in-person and remote users aware of each other's presence. Compared to existing systems such as Gather.Town that also use a map-based interface, HyWay's design is different in being skeuomorphic, with the virtual map mirroring the physical floor. This enables a natural mapping between the "locations" of the in-person users and that of the remote users, and transitions between such locations. To minimize the need for toggling between the map view (needed for neighbourhood awareness) and the call view, we overlay the call view on the active map and also provide a "neighbourhood view" showing the avatars of nearby users, thereby enabling the user to have neighbourhood awareness even while in the call view (See Figure 3).

3.4 Bridging the Agency Gap

Awareness naturally leads to the need for user agency, i.e., the ability of users to move between conversations as they wish. We wish to enable both in-person and remote users to enjoy such agency.

An in-person user who overhears another conversation or sees other users — in-person or remote — on the floor, could simply walk over to the other location to join that conversation. The only constraint is that interaction with remote users would be confined to the locations where the screens are installed.

A virtual user could do likewise by navigating their avatar on the map to get to the location of interest. However, in keeping with our goal of surprise-free interaction, such navigation, even though it is virtual, is subject to the constraints of the physical space. The speed of movement is akin to that of a in-person user who is actually walking on the floor (so there is no point-and-click teleporation in contrast to systems such as SpatialChat) and walls, furniture, etc. act as barriers that need to be navigated around. So, if a user sees (on the map) that another user is on the other end of the floor, they can be assured that it will be a while before the other user is within earshot and so can calibrate their discussion accordingly.

Porosity would ensure that a virtual user whose avatar is on the move would hear an ever-changing mix of "nearby" conversations as they traverse the floor. HyWay automatically juggles multiple audio streams under the covers to provide a natural and seamless experience, while also being resource efficient in terms of the number of audio streams subscribed to by the virtual user. We discuss agency further in Section 6.

4 RECIPROCITY

The goal here is to ensure that users who participate in HyWay can be seen and heard by others just as well as they can see and hear others. This would be akin to the physical setting where a user who steps into the hallway to mingle with others does so while being seen and heard by others and *not* by somehow hiding themselves.

The challenge is in ensuring reciprocity for remote users. Our overall approach is to meld technical enforcement with social constraints, which we discuss in detail next. Through technological enforcement, we ensure that (a) camera and microphone are used only when users (both in-person and remote) are in their camera's field of view (FOV), or in the HyWay zone (in the case of in-person users), (b) both remote and in-person users are made aware of each other's presence using face bubbles on the map, and (c) remote users should not be able to use an image or video proxy to circumvent reciprocity.

4.1 Enforcing *ON* State for Camera and Microphone

Unlike the typical audio-video conferencing applications, HyWay does not provide an on/off switch for the camera or the microphone. If a remote user wishes to get onto HyWay, they would need to have their camera and microphone turned on.

Since the remote user might be at home or another location, the user might need a way to filter out their surroundings and other activities happening, for them to be comfortable having their camera and microphone on. Our design leverages existing work on background filtering and noise suppression which is part of existing video conferencing system such as Teams [25].

Furthermore, just as an in-person user has the option to step away, say to answer a phone call in private, we provide remote users the option to put HyWay on hold to attend to their private activity. While on hold, the user can no longer see or hear any activity on HyWay and likewise they cannot be seen or heard either. However, the resources such as the calls that the user's client is part of are retained, so that rejoining after the hold period is much more lightweight than starting afresh.

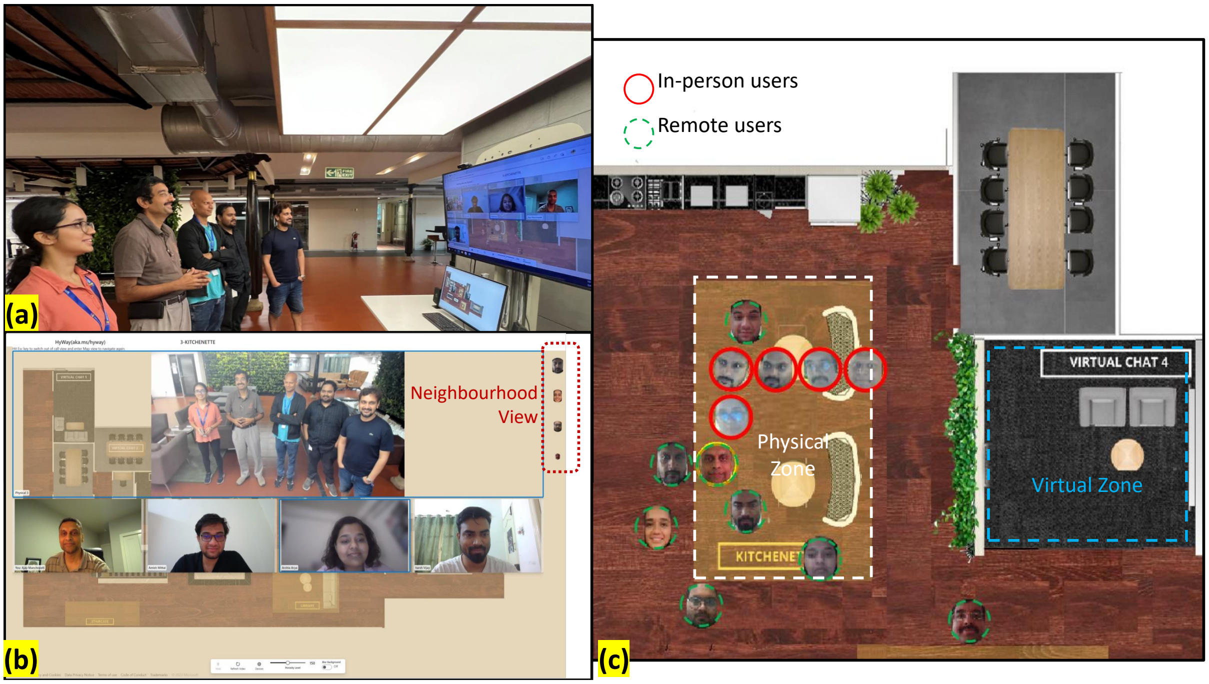


Fig. 3. (a) The physical floor view, (b) the remote user call panel view, and (c) the map view (with physical and virtual user bubbles marked).

4.2 Representing Remote and In-person Users through Face Bubbles

A remote user can be seen as an avatar on the HyWay map. However, there is no representation of an in-person user on HyWay. To maintain reciprocity between remote and in-person user, we represent the in-person users standing in the FOV of the camera as face bubbles on the HyWay map. To create the face bubbles, we use the face detection feature of Azure Cognitive Services Face API [10]. The Azure Face-Detect API generates face bounding boxes for the in-person users present in the FOV. We extract the faces from these bounding boxes and present them as face bubbles with a red halo on the HyWay map. Similarly, remote user avatars, captured during login, is also represented as face bubbles with green halo on the map (Figure 3(c)) — helping distinguish the two classes of users. Furthermore, to accommodate users who are colour blind, we make the halo of the remote users a dashed circle, as can be seen in the "map view" shown in Figure 3(c).

Finally, the face bubbles of users in the vicinity, with the bubble size as a function of the proximity, are also included in the "neighbourhood view" shown in Figure 3(b). This helps the user be aware of those in their vicinity even while they are busy conversing in the call panel and so not looking at the map.

4.3 Liveliness Detection

It is possible for a remote user to have their camera on and still not be in view. For instance, the camera's FOV could be occluded with a lens cap, or the camera could be pointed away from the user, or an enterprising user looking to game the system could even point the camera towards a picture (or video) of themselves. We term such attacks as spoof attacks.

Table 1. Liveliness Detection Experiment.

Input	Number of samples	Face-spoofing detector		Eye-mouth movement detector		Proposed Approach		Ground Truth	
		Spoof not detected	Spoof detected	Spoof not detected	Spoof detected	Spoof not detected	Spoof detected	Live	Spoof
Live Person	20	18	2	20	0	18	2	20	0
Photo of a person rendered on a tablet screen	20	15	5	1	19	1	19	0	20
Video-recording of a person rendered on a tablet screen	20	6	14	2	18	2	18	0	20
Person holding their face image	20	17	3	0	20	0	20	0	20

To detect and defeat such attempts, we employ a combination of technical enforcement and social constraints. The former is used where attempts at circumventing reciprocity might otherwise go unnoticed. HyWay uses liveliness detection to ensure that camera feed is of a live human face. Our liveliness detector involves a combination of face-spoofing detector [38], and eye-mouth movement detector. The face-spoofing detector extracts histograms corresponding to YCbCr [76] and CIE $L^*u^*v^*$ colour spaces [75] and uses an ExtraTreesClassifier [9] to distinguish a genuine face sample from a spoof attack. However, a user can game this detector by sending a genuine face image through a virtual camera device. To defeat such attempt, we use an eye-mouth movement detector on top of the spoof detector. To increase the robustness, we reject negative (spoofed) samples based on the spoof detector and check for eye and mouth movements on the positive (real) samples.

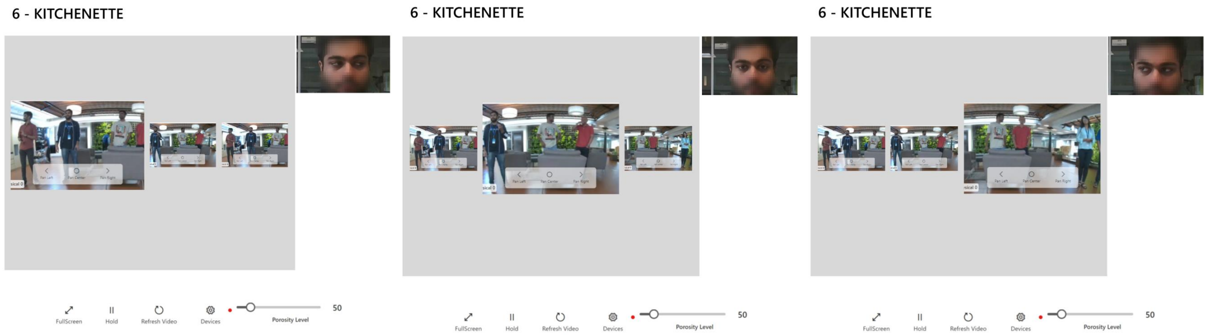
We evaluated our methodology using CASIA FASD [77] test dataset. This dataset comprises 30 subjects, with each subject containing a mixture of 3 genuine and 9 spoof attack video clips. We randomly sampled a subset of images and videos corresponding to the real people and their spoof attacks (photo/video recording of the corresponding person).

Table 1 shows the robustness of our approach (spoof detector + eye mouth movement detector) against various attacks. When a printed photo of a user is held in front of the camera, the spoof detector failed (i.e., it failed to detect the spoofed samples) 17 out 20 times. However, when the camera was instead pointed at the same photo but rendered on a tablet screen, the spoof detector did a little better (as the image quality was somewhat worse compared to the print) but still failed 15 out of 20 times. When the photo was replaced with a video on the same tablet screen, the failure rate dropped to just 6 out of 20 times. This is counter-intuitive, since we would have expected the video to look more "real" than a static photo and so the spoof detector should have failed more. We believe the decrease in the failure rate in the case of video is because of the poorer image quality compared to a static photo, which allows the detector to more easily identify spoofing. For all these cases, our proposed approach, which additionally incorporates the eye-mouth movement detector, is more effective in detecting spoofing, as reported in Table 1.

We rely on social constraints where technical means are challenging but social means are much more feasible. For instance, even our liveliness detection such as that outlined above can be gamed. The camera could be replaced with a virtual camera device that reads from a prerecorded or live stream video. However, in a social context, such attempts would be easily detected and frowned up by other users, thereby discouraging such behavior.

4.4 Adaptive Camera and Microphone Operation

The "presence" of the remote users is confined to the displays and their view of the physical space is determined by the FOV of the camera mounted on the display. So, there are limits to how well remote users can be seen and how well they can see, which, in turn, can pose challenges to reciprocity. For instance, in-person users who are standing behind a display may have no inkling that their conversations are being overhead by a remote user who is on the screen but staying silent. Likewise, two remote users who are talking to each other while the physical floor is seemingly empty might not be aware that an in-person user who is out of the view of the camera is listening. To mitigate such issues, we make the the operation of the camera, microphone and the speakers adaptive to the context of both the users and the physical setting.



(a) User looking to his left **(b)** User looking to his center **(c)** User looking to his right

Fig. 4. Peripheral awareness on HyWay for remote users based on their gaze change.

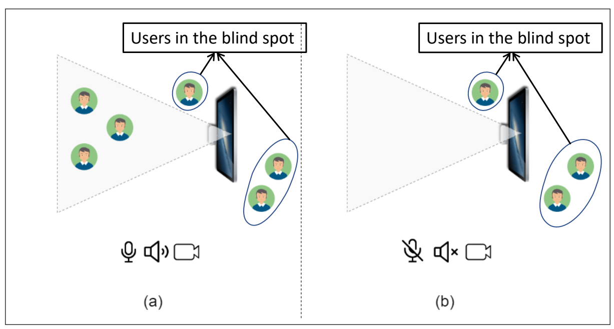


Fig. 5. Adaptive mute and un-mute on HyWay. (a) Mic and speaker are *ON* when HyWay camera detects in-person users (b) and *OFF* when no users are detected. All depicted persons are in-person users.

4.4.1 Peripheral Vision. The viewing angle of a modern display is very wide, e.g., 140-170^o for an OLED display [15]. However, the FOV of a webcam is much narrower, e.g., 65-90^o [5]. As a result, remote users on a display can often be seen by in-person users who are outside the FOV of the camera, which goes against principle of reciprocity. To bridge the awareness gap between the in-person and remote users, wide angle cameras are mounted onto the physical displays. This ensures that the FOV of the camera corresponds to the viewing angle of the display or the camera image can be trimmed to match the viewing angle of the display. On the client end, the camera stream is split into three horizontal overlapping segments of 90^o FOV each, as shown in Figure 4. The remote user's default view is the center tile, which is rendered as the largest tile, keeping the other two smaller. When the remote user scans to the left, HyWay detects this using the gaze tracker and expands the corresponding peripheral tile as shown in Figure 4. The gaze tracking algorithm uses the regression tree based face keypoints detector [52]. The coordinates of the pupils are then isolated from the face landmarks and compared to the center coordinate of the eye. The gaze is subsequently tracked based on the relative position of the pupils from the center of the eye.

4.4.2 Adaptive Muting and Unmuting. Unless we have a 360^o FOV of the camera and viewing angle of the display (e.g., a circular display) — both of which are hard to achieve with commodity hardware — there will be a blind

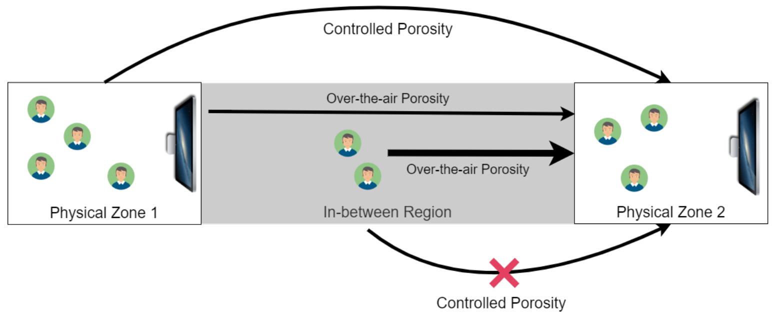


Fig. 6. Over-the-air versus controlled porosity. All depicted persons are in-person users.

spot (Figure 5), where in-person users and the remote users cannot see each other. Such a blind spot could create awkwardness noted above, where in-person or remote users might engage in sensitive conversation without being aware of the presence of others nearby.

To mitigate this issue, HyWay mutes or unmutes the speaker and microphone attached to the display on the physical floor depending on the context. Specifically, if no in-person users are seen in the vicinity of the display, we assume that the remote users, if any, on that display are engaged in conversation only amongst themselves. So, we mute both the speaker and the microphone to ensure the remote users conversation is not heard on the physical floor (e.g., by in-person users who are in the blind spot), and equally, the remote users are not able to overhear conversations among such in-person users.

However, when there are in-person users in the FOV of the camera, they and any remote users would likely be engaged in a conversation, which we believe would mitigate the above concern and hence the speaker and microphone could remain unmuted. Specifically, in-person users in the blind spot would likely become aware of the presence of the remote users and the in-person users in the FOV of camera could alert the remote users to the presence of in-person users in the blind spot, to avoid any awkwardness.

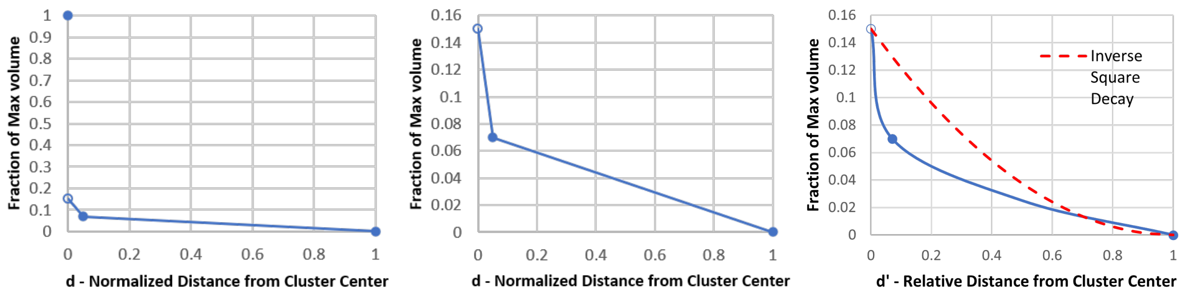
5 POROSITY

As discussed, porosity enables users, both physical and virtual, to overhear conversations in the vicinity. We start by explaining the need for "controlled porosity" over an alternative that we term as "over-the-air porosity", and then turn to the details of enabling controlled porosity.

5.1 Over-the-Air versus Controlled Porosity

Consider the setting depicted in Figure 6, with two physical conversation zones separated by hallway space that we call the "in-between" region. In theory, the voices of both the in-person and the remote users in zone 1 would be carried over the air to zone 2, where the microphone would pick up the sound and deliver it to the remote users in zone 2. However, there are multiple challenges with such (physics-based) over-the-air porosity.

First, the self-noise or equivalent input noise (EIN) [19] of the microphone in zone 2 could drown out any distant voice emanating from zone 1. For example, if the EIN of the microphone is 30 dB, a soft conversation (50-60 dB) would drop below the noise floor at a distance of 10-30 m and therefore be completely drowned out by the noise from the viewpoint of the remote users in zone 2. In comparison, the threshold of hearing for a (young and healthy) human is 0 dB [17] (corresponding to a sound pressure level (SPL) of 20 μPa [18]), which means that the same sound could still be heard by the in-person users in zone 2.



(a) Porosity Function (b) Porosity (Zoomed in) (c) Smoothed Porosity (Zoomed in)

Fig. 7. Porosity - Zone volume drop as a function of the relative distance from Zone.

Second, the speaker volume setting in zone 1, which determines how loud the voices emanating from the virtual users are, would attenuate the sound further if the volume is turned down, making it even more difficult for the zone 2 microphone to pick up the voices.

Third, even if the above technical issues were somehow circumvented, over-the-air porosity would mean that voices from not just the remote zones but also from the in-between region would be picked up. This could be problematic since unlike the in-person users who position themselves willingly in the vicinity of the zones (where we have a privacy notice displayed prominently), the users who happen to be traversing the in-between region would, in general, have not consented to participating in the system. So, we would not want the microphones to pick up the voices of these users.

In view of these challenges, HyWay employs (software-)controlled porosity, wherein the voices within each zone are picked up by just the local microphone and the audio streams from the various zones are mixed, after suitable attenuation, to create a composite stream that, in effect, incorporates porosity. The microphone sensitivity threshold is set so as to filter out voices from outside the zone. One of the advantages of such controlled porosity is that the user can control the extent of porosity that they would like, including, in the extreme case, opting for no porosity at all, as we explain below.

5.2 Calibrating Porosity

In theory, sound intensity (and hence the over-the-air porosity) drops off with distance as per the inverse square law [14], which suggests that controlled porosity should follow a similar trend. However, such a drop-off only holds in free space. In a space with obstructions, such as walls and furniture, the drop-off tends to be steeper. This calls for an empirical approach to calibrating porosity.

We conducted small-scale controlled experiments (elaborated on in Section 5.3), which yielded the following key observations:

(1) Users found it extremely difficult to focus on an ongoing conversation in the zone they were in if another conversation was overheard with a volume level that was greater 15% of the normal (i.e., within-zone) level. Note that the exact value varies with the number of zones with active conversations at that time (Section 5.3).
(2) When listening to background conversations, comprehensibility rapidly dropped with an increase in the number of active background conversations, with most users being unable to follow anything when there were 4 or more such active conversations (see Section 5.3). In comparison, in a physical setting, users can tune out such background chatter but this is far more difficult to do in a virtual setting.

We believe both of these differences arise because controlled porosity in HyWay does not provide an indication of the direction of the overheard conversation. In the absence of such a cue, users were unable to filter out background noise as effectively as they could in the physical setting.

The extent of these issues varies with the nature of the overheard audio source and from person to person. Therefore, in HyWay, we provide the user with a slider to control the degree of the porosity at an individual level. Based on these observations, the porosity function (Figure 7) is designed to have the following characteristics:

(1) As soon as the virtual user exits a physical zone, the volume of audio emanating from that zone is reduced to 15% of the normal volume. This leads to a disjoint drop at the beginning of the graph.
(2) The volume decay is extremely rapid at first followed by a more gradual delay. This ensures that far away zones are reduced to noise and do not overwhelm users.

In mathematical terms, we define a virtual user's normalized distance from a zone i's center (d_i) as:

$$d_i = \frac{\text{distance of the virtual user's avatar from zone } i}{max\_distance} \tag{1}$$

where $max\_distance$ refers to the greatest possible distance, i.e., the length of the map diagonal in Figure 9.

All the graphs in (Figure 7) show the decay in the audio volume of a zone with respect to d_i. As noted before, the volume starts at 15% (adjustable by the individual users) except when the user is inside the zone, which corresponds to $d_i = 0$. To let users listen to up to 4 conversations (beyond which comprehensibility becomes very difficult), we define a reference point $d_i^{background}$ that indicates the radius within which 4 nearby zones are covered. Note that $d_i^{background}$ would depend on the layout of the zones (i.e., how close or far apart these are), which might be different across locations.

Our aim, thus, is to allow users to hear audio from clusters that are beyond $d_i^{background}$ but not be able to make out the details of the conversation (so as to not overwhelm the users). In our small scale experiments, we observed that this effect is achieved when the volume is set to half the background volume level. As such, till $d_i^{background}$, the volume decays to half the initial background volume. For most map configurations, $d_i^{background}$ is a small fraction of $max\_distance$, so the initial volume decay is far more steep. Initial iterations of HyWay utilised a linear step function (Figure 7b), which was later replaced with a smoothed version (Figure 7c). For smoothing the volume, we utilised Bezier functions [4] as they provided the best experience in our small-scale experiments compared to power law and exponential decay. In addition, if there is an intervening wall between a zone and the virtual user's location, we scale down the zone's background volume by a constant factor of 0.3 (i.e., 70% volume reduction) to account for sound muffling by walls. We defer to future work a more careful model that takes into account both nature and the type of walls as, for instance, has been done for room acoustics simulation and virtual acoustic environments [54].

5.3 Controlled Empirical Study of Porosity

We conducted a controlled experiment to investigate the effectiveness of porosity and the impact of the number of zones with background conversations. Specifically, we wanted to understand how the number of conversations overheard impacts the optimal porosity level for the users.

5.3.1 Experiment Setup. Upon recruitment, the 10 subjects were each briefed about the HyWay system and the concept of porosity, and the intent of the experiment. After the initial briefing, each participant was put in a virtual conversation zone and was subjected to 2-3 scenarios with a varying number of nearby active background conversations. Participants were placed in each scenario for roughly 4-5 minutes and they participated in the experiment for a total of about 30 minutes. During each 4-5 minute scenario, the participant had to converse with another person in their own conversation zone, and were also asked to adjust porosity to the optimal level.

Table 2. Experiment Parameters.

Independent Variable	Dependent Variable	Control Variable	Confounding Variable
number of conversation zones [2,3,4]	porosity value [0-100]	number of people per conversation zone [2]	conversation bias & people bias

Specifically, they were asked to set the porosity level such that they could get a sense of the nearby conversations while being able to meaningfully engage within their own active conversation. This porosity level was recorded by the researcher upon completion of each scenario. The researcher reset the porosity level to 0 at the beginning of each scenario to minimize the learning effect and bias from previous scenarios. To avoid device variation, we ensured that the audio/video hardware in the physical zone and the machine configuration of the users participating virtually was identical. Each user was provided with a Lenovo Thinkpad X1 Yoga laptop with an i7 processor and 16GB of RAM during the experiment.

5.3.2 Study Parameters. As shown in Table 2, we had 1 *independent variable – number of background conversation zones* (2, 3, 4), and 1 *dependent variable – porosity level* [0, 100], which means that the porosity level can be set to anywhere between 0% (zero volume, so background conversations are not heard at all) to 100% (such conversations are heard without being attenuated). We controlled the number of people per zone to 2, and let the participants talk on a topic replicating natural conversations. We acknowledge that there might be bias related to the topic of conversation, which might have impacted our participant's perceived optimal porosity level.

5.3.3 Data Analysis. We had a total of 29 data points in the format (x, y), where x represents the number of active background conversation zones that can range from 2 to 4, and y represents the optimal porosity level as adjusted by the participant for the chosen number of background conversation zones. We normalized all of our values, and then applied the Pearson correlation test, which is a measure of linear correlation between two sets of data [35]. We got $r = -0.2675$. Negative r value in our context implies that the number of conversation zones and optimal porosity level are negatively correlated. This further demonstrates that, with the increase in number of conversation zones, the optimal porosity level for the end user will decrease. Furthermore, we passed all of our data points into a Decision Tree Regressor [63] to get an idea of the average optimal porosity level for each configuration. As observed in Figure 8, the optimal porosity level regressed progressively decreases with an increase in the number of background conversation zones, which is consistent with the above correlation results.

6 AGENCY IN USER MOBILITY

In a physical hallway setting, users have the ability to move around and join new conversations as they wish. Indeed, such *agency* in mobility is a natural consequence of awareness, which has been a key design focus of HyWay. In-person users on HyWay can simply move around — either to a chosen physical zone to engage in a new conversation or away from all zones — just as they would in a purely physical hallway setting. So, we focus here on providing remote users with similar agency.

To enable remote users on HyWay to navigate from one conversation to another, we have designed and developed a map-based interface (see Figure 3 and Figure 9) that allows remote users to seamlessly navigate to different conversations happening in the hallway, just as they would in the physical setting.

To move from one location to another, a remote user would move their avatar on the map using their cursor keys. This ensures gradual movement unlike a point-and-click interface, such as teleportation in SpatialChat [23]. This helps ensure a surprise-free experience on HyWay, where users have the assurance that others cannot appear unexpectedly.

We designed a skeuomorphic map-based representation of the real-world physical space. This is a key distinction from systems such as Gather.Town [13] that also use a map-based interface, which is typically synthetic. The

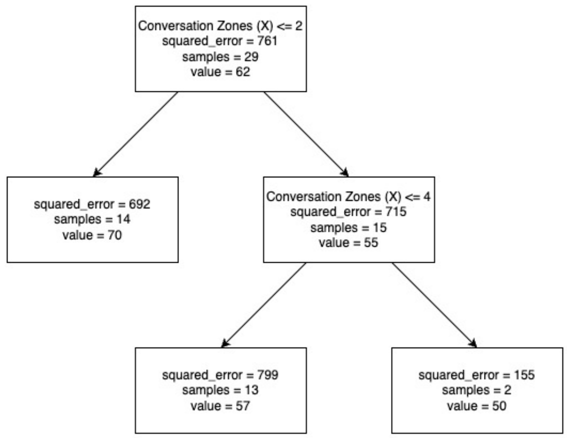


Fig. 8. Decision Tree obtained from the data points collected during the experiment. As clear from the tree, the regressed optimum porosity value regressed decreases with an increase in the number of conversation zones. The high error values is recorded because participants chose high difference values when changing porosity as the number of active calls increased.

correspondence of the map in HyWay to the physical space ensures consistency of experience across in-person and remote users. For instance, the set of conversations in the "vicinity" that a remote user in a zone overhears through porosity approximates what an in-person user in the same zone would overhear. Furthermore, the direct correspondence between locations on the map and those in the physical hallway means that in-person and remote users can move to the "same" location if they so wish. For instance, if an in-person (or remote) user moves from one zone to another, a remote (or in-person) user they were conversing with could do the same and thereby continue their conversation in the new location.

The map in HyWay consists of three object layers: (1) the bottom layer depicts the floor texture (e.g., wooden, carpeted, marble), (2) the middle layer consists of items such as furniture (e.g., tables, chairs), and indoor plants to give a realistic feel to the user, and (3) the final layer consists of the user avatars (face bubbles) representing remote and in-person users (see Figure 3).

We define collision boundaries on the map layers, which restricts the user avatars from stepping over objects (such as furniture), navigating through the walls, and going beyond the defined boundaries of the space. Such restrictions help maintain parity in the movement of remote users and in-person users, thereby contributing to the surprise-free operation of the system, e.g., remote user cannot cut through walls and furniture and show up in a new location faster than in-person users making the same trek.

We define conversation zones on the map, where users can congregate and have conversations. A physical zone is one where there is a display on the physical floor and so in-person users and remote users can mingle with each other, while a virtual zone is one where there is no display on the physical floor and hence remote users alone can get together. The map in Figure 3 shows both physical zone ("kitchenette"), and a virtual zone. Porosity ensures that users in virtual zones can still overhear and be overheard by nearby in-person and remote users. Since all users, both in-person and remote, are depicted on the map, users are aware of who is around and so could overhear conversations, much like in an all-physical setting.

In our current implementation, remote users are unmuted only when their avatar is within the confines of a conversation zone. When they are on the floor and outside all the conversation zones, they remain muted but can overhear muffled conversations from the zones in their vicinity. Although in principle we could have the remote users be unmuted at all times, leaving them muted except while in the conversation zones helps reduce porosity clutter (as pointed out in Section 5.2, users disliked overhearing too many background conversations), besides simplifying the task of dynamic call management (Section 6.2). Next, we discuss two technical challenges that arise from the agency that remote users have in moving about in the floor: calibration of the speed of movement and dynamic call management.

6.1 Auto-Calibrating Speed of Remote Users

While the design aspects discussed above (e.g., a map layout that reflects the physical layout, collision boundaries that confine remote user avatars to the paths that in-person users can take) help with parity in the movement trajectories of the in-person and remote users, and thereby avoid surprises, there is another key element: the speed of movement of the remote user avatars. Ideally, the speed of movement of the remote user avatars should be consistent with the speed of in-person users on the floor, so that the time taken for both classes of users to get from say Zone 1 to Zone 2 on the floor is similar.

Having the speed of remote users match that of in-person users is challenging in general. For instance, the speed of movement of the latter might be dependent on the context, e.g., slower when a user needs to weave through a crowded conference hallway during the coffee break than when a user is walking through a sparsely populated office floor. Therefore, rather than configuring a fixed speed, which might be invalid in some or many settings, in HyWay we devise a simple procedure to *auto-calibrate* the speed. This would automatically help ensure that the speed is appropriate for the setting on hand.

Our auto-calibration procedure is based on the sightings of the same in-person user in different physical zones over time. We detect and match faces using the Face - Verify API [10]. [4] Say, an in-person user is last seen in Zone 1 at time t_1 and then next seen in Zone 2 at time t_2. $\Delta t = t_2 - t_1$ then provides an upper bound on the travel time for the in-person user between zones Zone 1 and Zone 2. However, there are many possibilities: users might not take the direct path from Zone 1 to Zone 2 or they might pause on the way (say, to chat with someone), or different users might walk at different speeds.

To overcome these issues, HyWay measures travel time for multiple users and multiple walks by each user to obtain a distribution like the one pictured in Figure 10. Then, we pick a low percentile, specifically the 10^{th} percentile, as our estimate of the travel time from Zone 1 to Zone 2 along the direct path and with no stops. [5] To verify our estimate of travel time, we also perform a controlled experiment where we ask a set of users to walk directly from Zone 1 to Zone 2 at a pace that is normal for them. Since there would be a spread of speeds across the users, we take the median (50^{th} percentile) of this set of recorded times to be the "ground truth" of the direct travel time. As shown in Figure 10, the travel time estimated by auto-calibration is a good match for the "ground truth" obtained by having a controlled set of users walk directly from Zone 1 to Zone 2.

After estimating the travel time between Zone 1 = Library and Zone 2 = Kitchenette (Figure 9), we estimate the speed of the remote user's avatar as the in-game distance between Zones 1 and 2 (i.e., 68.5 units, where a "unit" represents the smallest updatable distance on the map, which is set by the game layer (Section 7), and corresponds to 0.26 meters on the physical floor) divided by the estimated travel time (15.00 seconds), yielding a speed of 4.56 in-game units per second. This is the speed at which the avatar would move if the remote user keeps a cursor key pressed continuously.

Finally, HyWay at present only allows user avatars to move in one of four directions — up, down, left, or right, corresponding to the four cursor keys pressed individually. Therefore, the in-game distance between two locations is computed with this constraint in place and the avatar speed calibrated accordingly. If we were to augment HyWay to allow more flexible movement (e.g., diagonal movement, say by pressing two cursor keys

[4]Note that such face *matching* is distinct from face recognition.
[5]We avoid the minimum because it is susceptible to spurious face sighting and matching

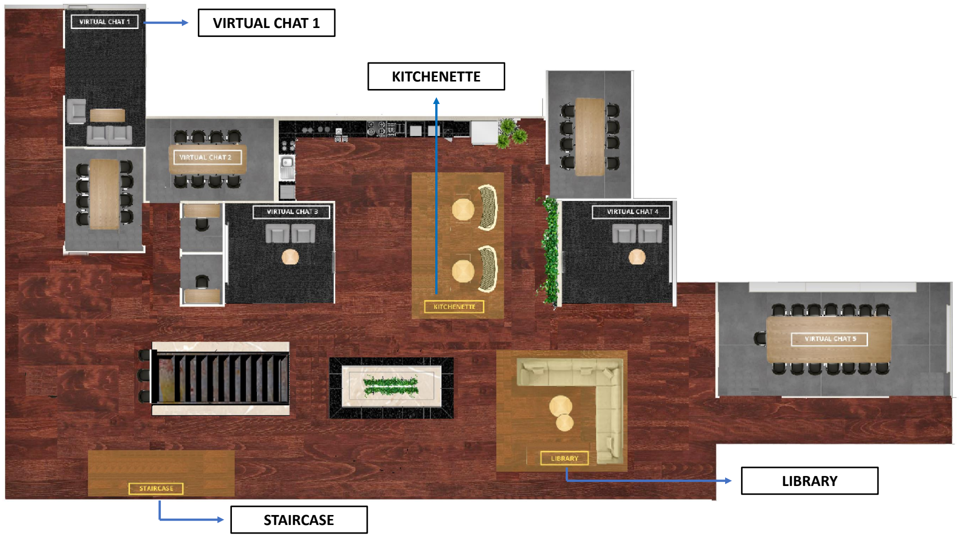


Fig. 9. HyWay map for the office workplace from Figure 1. Staircase, Library, and Kitchenette represent the physical zones. The map also includes virtual-only conversation zones.

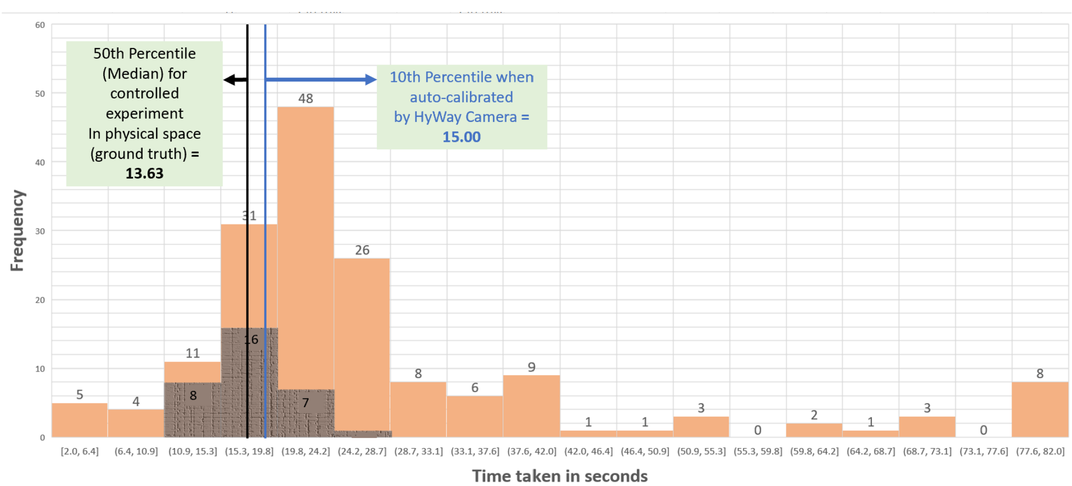


Fig. 10. Histogram representing the distribution of time taken for multiple walks from one zone to another. The orange distribution shows the time recorded by HyWay camera during deployment. There are 169 such samples with 10^{th} percentile as 15 seconds. The grey distribution shows the time taken by users in our controlled experiment. There are 32 such samples with a median of 13.63 seconds.

simultaneously), the in-game distance would have to be computed accordingly. The rest of the methodology to auto-calibrate speed would carry over as is.

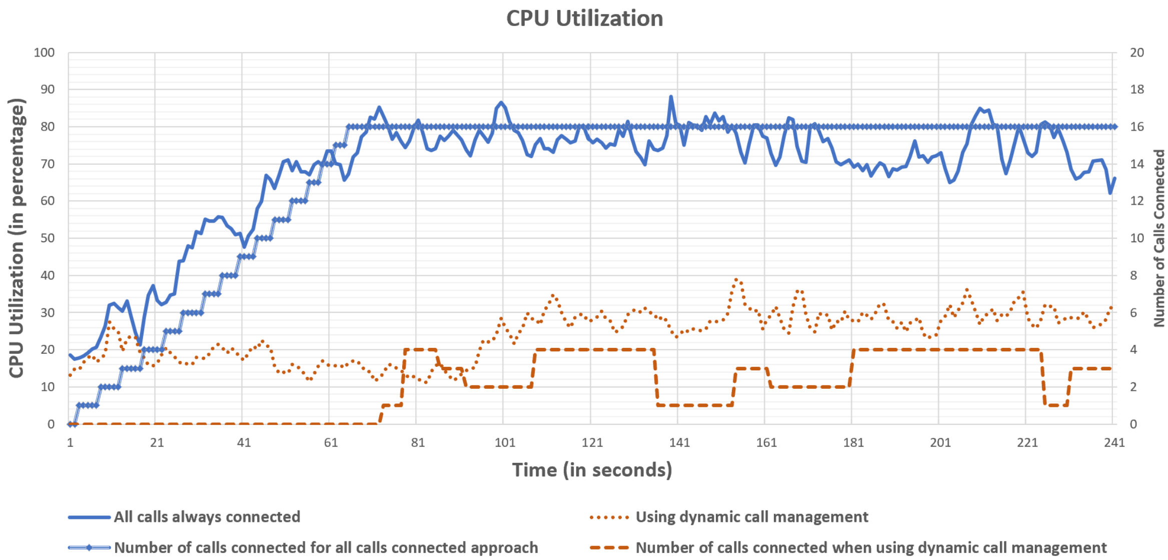


Fig. 11. CPU usage of browser thread and audio service when recorded over the deployed system for 16 active zones. For the all-calls-always-connect approach, time 1-80 seconds indicate the increase in CPU usage as the client attempts to connect to all the 16 zones, sequentially. The red line represents a much lower CPU usage using dynamic call management approach. Under the dynamic call management, time 1-80 represents when the remote user does not move and stays at a location which is far from all zones. The time stamps after 60 seconds represent a random walk around the map simulating normal user behaviour. Client runs on system with an Intel Core i7-8650U Quad-core at 1.90 GHz, 16 GB RAM at 1867 MHz. The experiment was done with 16 zones to show distinctive difference between the two approaches.

6.2 Dynamic Call Management

As the remote users move around, the ambient sounds they hear also change, just as it would happen in a physical setting. Under the covers, HyWay has each remote user's client join multiple calls, one corresponding to each zone, as we elaborate on in Section 7. The question is how the user client should manage these multiple calls.

Our initial approach in HyWay was simply to have each client join the calls corresponding to all the zones right up front at initialization time, before the remote user's avatar is allowed to join any conversation zone. While being simple, this all-calls-always-connected approach has three downsides: first, increased startup time because of the need for the client to join all the calls up front (which needs to happen synchronously by acquiring locks on the main thread, for the reasons noted in Section 7); second, increased CPU load on the client; and third, increased demand on network bandwidth. All of these issues are exacerbated when the number of zones is large. For instance, in Figure 11, with a total of 16 zones, the initialization phase lasts over 80 seconds and the CPU utilization rises to 80%, making the client sluggish.

As an alternative, we devised a dynamic call management approach, wherein each client only joins the calls corresponding to the zone it is in and the zones in the vicinity. The zones in the vicinity are defined by two factors. The first is the porosity neighbourhood, i.e., the zones (up to 4, see Section 5) which are in the vicinity of the remote user on the map. The second arises from the speed of the movement of a remote user's avatar versus the time taken to join a new call. The faster that the avatar can reach a new zone, the more in advance the joining of the call corresponding to that zone must be initiated. It takes 300ms to connect to or disconnect from a call, so multiplying this time with the speed of avatar movement (4.56 units/second) we obtain a trigger distance of

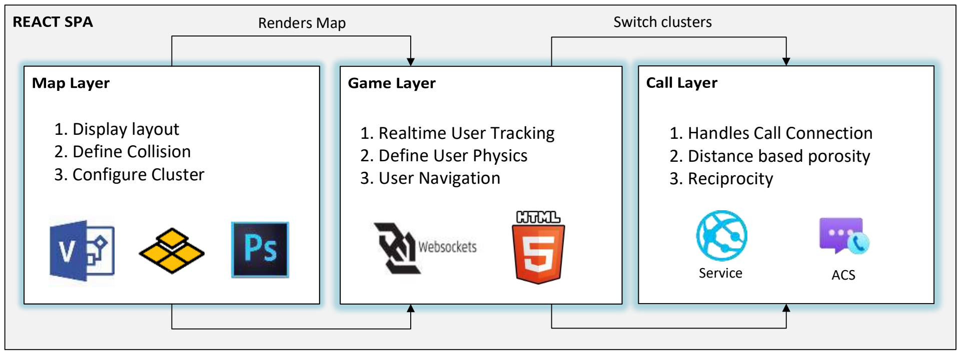


Fig. 12. Three layered architecture of HyWay.

1.37 tiles from a zone for the client to initiate the process of joining the corresponding call. This helps ensure a seamless user experience.

Accordingly, with dynamic call management, as the remote user moves around the map, we connect to the zone(s) which the user is approaching and disconnect from the zone(s) which are now far from the user. The benefits can be seen in Figure 11, where long startup delay and the high CPU load are both avoided. To make it easy to visually compare the all-calls-always-connected and dynamic call management curves, we initiate call connections in the latter case only after the 80 second mark, which corresponds to when the client has connected to all calls in the former case.

In summary, dynamic call management helps in drastically reducing the CPU utilization and making it independent of the total number of zones on the map. With this approach we can now scale the map and the client to support a large number of zones.

7 IMPLEMENTATION

We separated HyWay logically into three main layers (i) Map layer, (ii) Game Layer, and (iii) Call layer, as shown in Figure 12. The idea here is to decouple components in a way that can support multiple maps and allow anyone looking to deploy an instance of HyWay to define (a) the game physics, e.g., navigation rules such as the allowed movement directions (e.g., 4 (up/down/left/right) or 8 (additionally including the diagonal directions)), using cursors along with keys, etc., and (b) custom properties such as collision, wall muffling for porosity, etc.

The first layer is the Map layer, where users define a virtual representation of the physical space and its properties to make it realistic. Here we used tools such as Microsoft Visio [29] and Adobe Photoshop [2] to create a realistic virtual representation of the space. Layered on top of this, we have the Game layer that has been implemented using an `HTML 5` based game engine — Phaser3 [20] — which helps users navigate the space defined in the map layer. The topmost layer is the Calling layer, where we leverage Azure Communication Services (ACS) [3], which is set of communication APIs (e.g. calling APIs, video APIs etc.) to enable on-the-fly communication between users.

Azure Communication Services (ACS). In HyWay, we require custom calling experiences for both the remote and the in-person users. ACS provides a set of `APIs` to enable Internet-based audio/video calling and provides UI libraries to create custom layouts and experiences. To create a space with multiple concurrent conversations that gives HyWay users the ability to overhear nearby conversations, we start multiple calls and suitably attenuated the audio from these calls to provide users a distance-based porosity experience. Each zone has an associated call identified with a unique group ID, which enables everyone in a particular zone to join the same call and interact with each other. We also create custom galleries, for instance, to have a larger video tile for the in-person users who are on the physical floor (see Figure 3).

User Client. We built a React-based single-page application with Fluent UI [12] and Phaser 3 [20] to load a top-down, RPG (Role-Playing-Game) style browser game interacting with ACS calling components where each avatar represents the user in virtual space. The Phaser GameScene embeds a map mirroring the physical floor. We started with the floor map of the physical space which allowed us to understand the boundaries, collision objects and scale of the selected space. Next, we added textures to this floor map using Photoshop. The top view of objects such as furniture and plants, representing their counterparts on the physical floor, were added on the map (see Figure 3).

In the next step, we defined collision boundaries based on object locations by dividing the map into tiles in the Tiled [27] app. Phaser uses this collision configuration to ensure that user avatars on the map do not cross collision boundaries and only move within the allowed spaces.

The game map allows the physical users to get a sense of the location of the virtual users. It also allows the virtual users to navigate the map and participate in different conversations happening on the floor by leveraging the porous nature of the calls, as discussed in Section 5. When a user client connects to HyWay, we connect the user to calls corresponding to multiple conversation zones in the vicinity, as explained in Section 6.2. A virtual user can take part in a call by moving to the location of the conversation and hitting the Space key to open up the call panel. Client-side implementation of HyWay comprised 2981 lines of code (LoC), excluding `json` configs, map binaries and standard node dependencies.

HyWay server. We have developed a Node.js-based server that keeps track of the current locations of different virtual users. When new users join HyWay, the client initiates a socket connection with the server. The server then communicates the current state of the users, i.e., the locations of the users across the call zones, to the client app. The client app uses this state information to display user avatars on the map. When a user navigates through the map, location updates are sent to the server, which broadcasts these to the other clients. The clients then update the remote-user avatar's position on the map The server-side implementation of HyWay comprised 78 LoC (excluding `json` configs and standard node dependencies.)

HyWay Inference server. We built an inference server using a python-based web-framework called FastAPI [11]. We have trained models for liveliness and gaze detection, and for defishing, i.e., reversing the distortion in a fish eye image (see Sections 4.3 and 4.4.1). These have been hosted as REST APIs, which take images as input and return an inference. Each client extracts frames from the call video streams or from the local camera, and invokes these APIs at regular intervals.

8 USABILITY STUDY

Since early 2022, we have deployed HyWay at 5 events, with an intent to put our conviction to test and gather user feedback that has helped shape subsequent iterations. These deployments have helped us better understand the relationship between the familiarity among the participating users, the topic of discussion, and the layout of the space to concepts such as porosity, reciprocity, and agency. Table 3 summarizes the various HyWay deployments. [6] For the first four deployments (Section 8.1), we examined user experience through surveys (27 respondents in all). In the *Event 5* deployment (Section 8.2), we supplemented the survey (29 respondents) with interviews with a subset of the participants (10).

[6]The deployments reported in this paper were approved based on a privacy review. Furthermore, the survey form and methodology used in our deployments and the controlled user study have been approved by our IRB.

Table 3. Various deployments and the nature (facets) of these events.

Event	Space Familiarity	People Familiarity	Topical Connection	Duration	Engagement format	Audience size & mix
Country A, Event 1	Medium to High	Medium to High	Medium	3 days, 4 hours/day	Demos, Coffee breaks	~ 15 in person, ~ 65 remote
Country B, Event 2	Low	Low to Medium	High	3 hours	Demos, Coffee breaks	~ 20 in person, ~ 40 remote
Country B, Event 3	High	Medium	Low	4 hours	Informal	~ 100 in person, ~ 25 remote
Country B, Event 4	Medium	High	Low	3 hours	Informal	~ 10 in person, ~ 7 remote
Country B, Event 5	Low	Low	High	‘ 2 sessions, 3 hours/session	Demos	~ 50 in person, ~ 70 remote

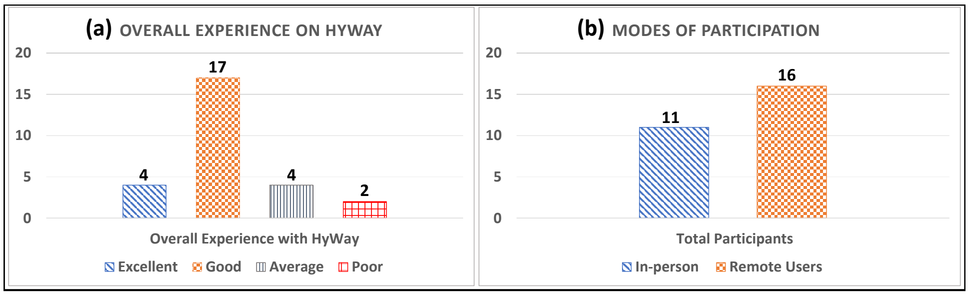


Fig. 13. (a) Overall HyWay experience. (b) Participation mix.

Note that our focus on events means that these deployments were confined in space and time, thereby mitigating concerns pertaining to privacy that might arise in the context of always-on 24x7 deployments. We defer exploration of the latter to future work.

Participants. Participants in all the deployments belonged to the same organization, Microsoft. Email invitations were sent to these participants, describing the nature of the setup and a brief tutorial on how to use the system. In all the events, the participants generally had no prior knowledge or experience of using HyWay. However, they were experienced in using existing video conferencing services such as Teams [25] and Zoom [30]. Some participants were aware of services such as Gather.Town [13] and AltspaceVR [31]; however, only a few knew how these services worked. Participants were free to use the system and *no* incentive was provided in any of our deployments.

Event space and setup. *Event 1* hosted series of talks separated by coffee breaks. Three HyWay zones were deployed during these breaks to enable informal mingling between the in-person and remote users. In addition, HyWay was used to host 16 demos spread across the floor map. The audience navigated the floor map and joined different demo sessions, which they learnt about either from the demo titles or by overhearing nearby presentations through porosity. *Event 2* also hosted a cohort of users who presented their work in a series of talks. HyWay was again used to enable informal mingling amongst both in-person and remote users during the coffee breaks. *Event 3* was a social event in a building atrium, where in-person attendees mingled with one another over wine and appetisers. Three HyWay conversation zones were set up to enable remote users to join in. *Event 4* brought together a small cohort of users representing different groups of the organization, celebrating their success over wine and snacks. Two HyWay conversation zones were set up to include attendees who could not make it in person. *Event 5* was a large hybrid poster event where we had both in-person and remote attendees engaging with both in-person and remote presenters.

8.1 Results from Initial Deployments (Event 1 — Event 4).

At the end of each event, we reached out to the participants by email and asked them to share their experience with HyWay via an anonymous online survey. The survey comprised 11 questions for the in-person participants

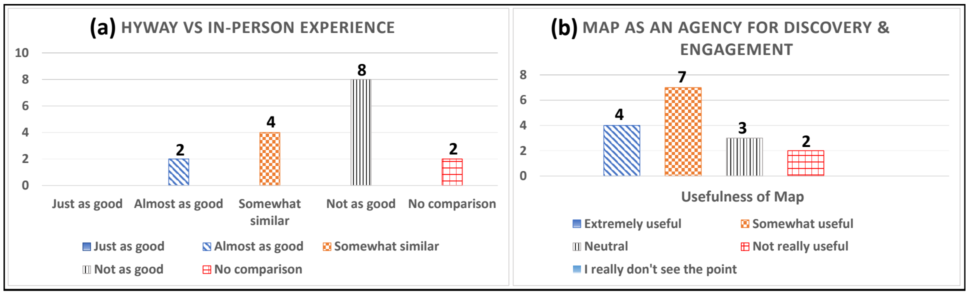


Fig. 14. (a) Remote users' experience as compared to in-person experience. (b) Benefits of floor map as an agency.

and 16 questions for the remote participants. Some of these questions provided checkboxes to allow selecting more than one choice while others used radio buttons to only allow a single choice. The estimated time to complete the survey was 5-7 minutes. In total, we received 27 responses. Participants also provided their descriptive opinions in the comments section of the survey.

Overall, the participants' response has been positive, as shown in Figure 13. Virtual user, V_7, [7] from Event 2 said *"I had a feeling of social interaction with people new to me in a way that was much more like in-person meetings versus just showing up on a Teams call. It was invigorating and engaging, and I really enjoyed the engagement ..."*, and V_{10} said *"It creates a shared/common space for virtual and remote team members. It seems great for serendipitous chats ..."*. Even participants who voted *"poor"* indicated that they had a better hybrid experience when compared to an all-virtual setup experienced on other video conferencing platforms. For instance, V_6 from Event 2 mentioned that the *"Combination of remote and in-person conversations (when they worked) were much better than in Teams."*. However, at times the system crashed unexpectedly, *"and long load time to restart [HyWay] meant that I missed large chunk of poster session."*

While the overall experience of using HyWay was positive, it also became clear that this is just the first step and there is more work to be done before a remote user can find the experience to be as good as attending in person, as indicated in Figure 14(a). Some of the descriptive feedback was about improving the UI fidelity. For example, V_8 from Event 3 said, *"The UI is of low fidelity. I don't have an avatar to represent myself, just a square (sprite) with my name on it."* To address this, we introduced (from Event 4 onwards) face bubbles to represent in-person and remote participants on the map. This provided users the ability to recognize others and thereby engage more effectively (as we touch on in Section 8.2).

There were instances when conversations among remote users unintentionally spilled over into the physical realm and disturbed physical users. For example, V_1 from poster Event 2 mentioned that *"I had some conversations with only remote participants at an 'empty' station where no one was physically present. We didn't realize that our conversation was loud enough to disturb people at a nearby station. I'm not sure what would have helped in that case, maybe a way to control volume at the physical station?"* Concerns such as this motivate augmentation of HyWay with adaptive mute/unmute (Section 4.4.2).

In all our deployments, we have used a map interface to provide users, both remote and in-person, awareness of who else is around, thereby enabling them to discover people and engage with them. As shown in Figure 14(b), the majority of respondents found the map to be useful. However, Events 3 and 4 were held in compact, single-room spaces, so some remote participants found the map interface to be not so relevant.

We observed that events where users have strong familiarity with the space or with other participants were the ones where the skeuomorphic representation of space was found to be particularly useful. Many participants

[7] *V* and *P* stands for remote (virtual) and in-person (physical) participants.

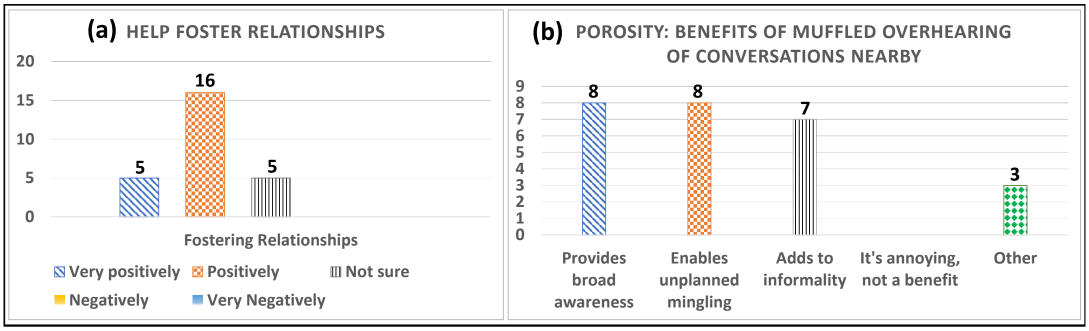


Fig. 15. (a) HyWay's ability to foster relationship. (b) Benefits of muffled conversation.

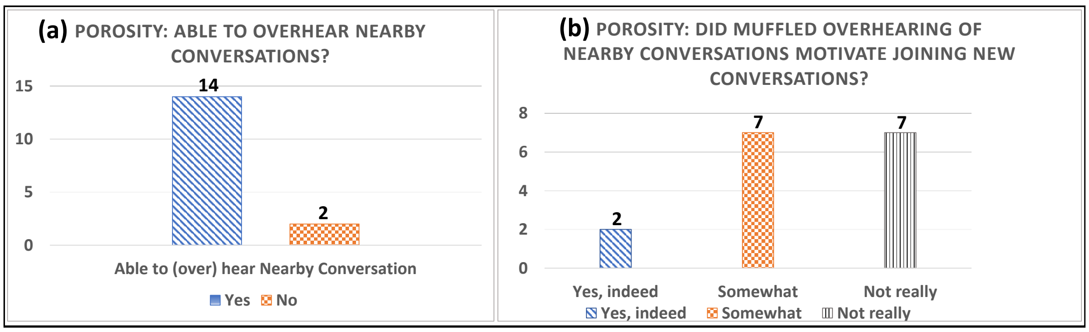


Fig. 16. (a) Ability to hear muffled conversations. (b) Motivation to join new conversations.

shared similar opinion as participant P_{13} from Event 3, who said *"Being able to chat with folks joining remotely while also interacting with people in the physical space was nice and being able to switch between the various stations (while providing the same capability to the remote attendees) was a good bonus."*

It is very natural for an in-person attendee to experience the buzz and the energy of an event. The porous nature of the environment not only facilitates experiencing ambient noise (buzz) but also provides the ability to overhear nearby conversations, thus enabling seamless transition between conversations. We believe *porosity* is one of the key tenets to enable remote attendees to have as close to an in-person experience as possible. Our survey results captured a similar sentiment (Figure 15(b)). It was promising to observe that this was not perceived as an annoyance or distraction. Some of the attendees explicitly called out moments where they recognized the voice from a nearby conversation and reached out to engage. Participant V_9 in Event 4 stated, *"I love the concept of the ambient audio, so that people virtually can tune in to things happening around them, not just the active conversation. The visualization of the space and movement between different areas is super interesting to for when you are remote."* Similarly, participant V_4 in Event 4 stated, *"Loved the informality; porosity made it feel much more real; spatial layout/mapping of the building/space was super cool!"*.

Most remote participants were able to overhear nearby conversations when using HyWay (Figure 16(a)), yet we found many not finding it sufficient to motivate joining new conversations (Figure 16(b)). As we take a closer look at the results, we again observe that strong ties (familiarity) with people play an important role in the perceived value. Most of these remote users who responded "not really" were attending events where their familiarity with other participants was low.

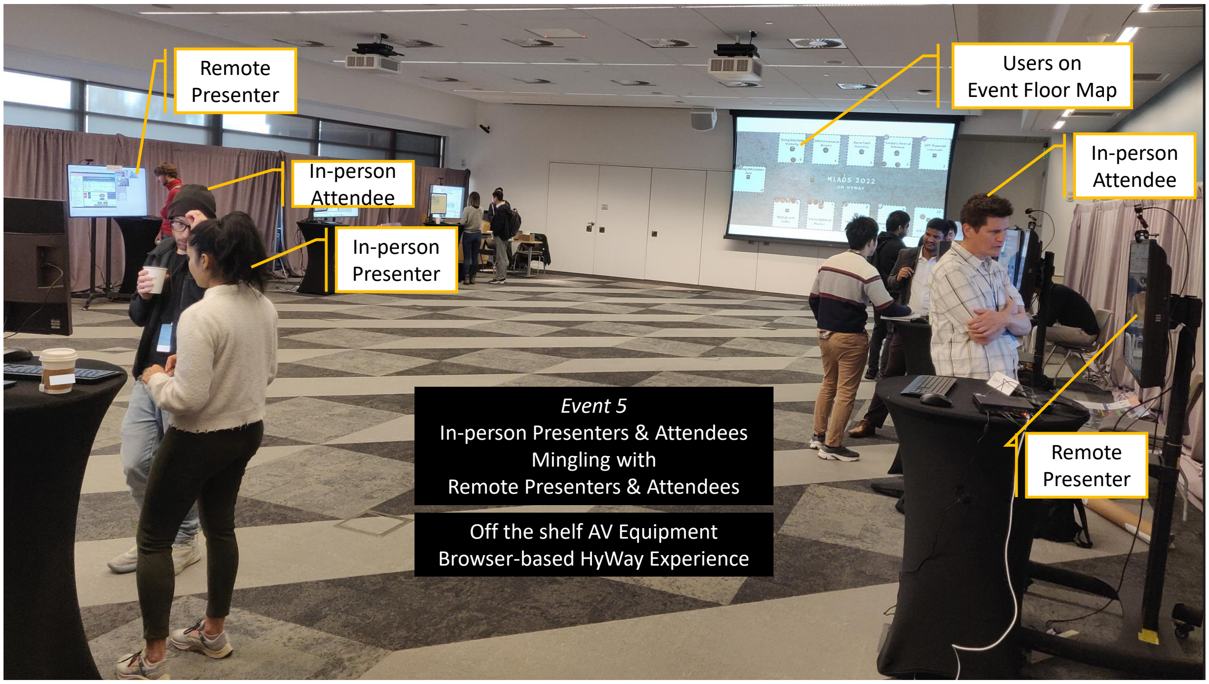


Fig. 17. HyWay setup during the hybrid poster event (Event 5). The adjacent distance between each booth was ~ 7 ft.

This is probably an expected outcome even for an in-person attendee with weak-ties with other participants. In fact, in-person participant P_4 in Event 3 (which involved informal mingling in a social setting) stated, *"... I was able to say 'hey [Name], what are your thoughts', but I don't think I ever even introduced myself, so those on the call didn't know who we were ... "*. In the future, we intend to explore alternate nudge techniques that might make it easier for remote users to join and participate in conversations with less familiar people, making it an experience that is perhaps better and beyond what is possible for an in-person attendee. At a broader level, we explored whether such a system can help foster relationships between people who are geographically separated, and we have had encouraging results, as shown in Figure 15(a).

8.2 In-depth User Feedback from Event 5

Event 5 was a large hybrid poster presentation event, where users presented their work either remotely or in-person (Figure 17). Email invitations were sent to all the participants, describing the nature of the setup and a brief tutorial on how to use the system.

The survey results from the previous deployments only give us a high level understanding of the participants' experience with the HyWay system. In the case of Event 5, we additionally conducted in-depth 1:1 interviews with a subset of the participants (Table 4) who had completed the survey and volunteered to be part of the study (10 interviewees out of the 29 survey respondents). These interviews allowed us to probe deeper into the participants' experience during the event. No incentives were provided to the participants for their participation in the interview. All interviews were recorded, transcribed, and reviewed. After open coding of the transcripts (breaking participants' response into discrete parts, i.e., codes), we used axial coding (drawing connections between the codes) to examine user responses. [71].

Table 4. Interview participants' demographics and their current work status.

ID	Gender	Experience with any social VR platforms	Work type
RD_1 *(Remote Demoer)*	M	No	Permanent work-from-home
RD_2 *(Remote Demoer)*	M	No	Permanent work-from-home
PD_1 *(Physical Demoer)*	F	No	Hybrid
PD_2 *(Physical Demoer)*	F	No	Hybrid
PD_3 *(Physical Demoer)*	M	No	Hybrid
VA_1 *(Virtual Attendee)*	M	No	Permanent work-from-home
VA_2 *(Virtual Attendee)*	M	No	Hybrid
VA_3 *(Virtual Attendee)*	M	No	Hybrid
EO_1 *(Event Organizer)*	M	No	Hybrid
EO_2 *(Event Organizer)*	M	Yes	Hybrid

Findings. We present our findings from: (1) the participants' high-level feedback from the survey and (2) learnings from their subsequent in-depth interviews.

Survey results from the event were similar to the results from our prior deployments. In total, we received 29 responses (24 remote and 5 in-person participants). Of these, 4 rated their experience as "excellent", 10 as "good", 8 as "average", and 7 as "poor". Based on the descriptive comments in the survey, the respondents who rated their experience as "poor" reported technical issues (e.g., audio/video equipment issues or UI flickering) as their predominant concern.

Map as an enabler for agency — discovery and engagement — also received positive responses, with 20 (out of 24) remote participants voting the map-based interface to be either "extremely useful" or "somewhat useful" in discovering poster zones on HyWay. Describing the experience in the comments section of the survey, remote participant V_2 stated, *"It [map interface] allowed people to move in and out of the conversation in a more fluid way than you can in video conferencing platforms."* Also, 21 (out of 24) remote participants liked the porosity feature. However, a few participants, who identified themselves as remote demoers, reported dialing down the porosity level to have a more focused discussion during the poster event.

Participants' feedback from the interviews.[8] Most participants (Table 4) mentioned this was a first-time experience for them at a hybrid poster event. When asked about their experience with other virtual interaction platforms such as Gather.Town [13], most of the participants indicated that they never had experience with such platforms. However, all the participants had the experience of attending in-person poster events.

Almost all of the participants in our interview study have adopted a hybrid work style, except for 3, who have a full-time WFH arrangement. They all mentioned that they miss out on the short, serendipitous conversations that used to happen in the hallways, by the water cooler, and the kitchenette areas. A few participants, who described themselves as extroverted, came up with ways to recreate those unstructured conversations with their colleagues. For instance, remote attendee VA_2, who is a manager, stated, *"I used to ping people and say 'Hi, what's going on?', whenever I saw their status 'available' on Teams chat."*

Our post-deployment interview centered on 5 main themes, covering the important aspects of our system design, implementation, and user experience:

(a) Informative and easy-to-use 2-D map interface. All participants found the 2-D map-based interface to be informative and easy to navigate. One of the remote participants, VA_1, mentioned that the map provided a *"bird's eye floor plan"* that helped him look over different poster booths. He added that *"... it was kind of a bird's eye floor plan. I thought that was pretty cool ... I had the option of kind of looking at all the titles of the rooms ... so I could kind of plan out a little bit which ones I was looking at."*

[8]The interview responses have been transcribed verbatim, without fixing the grammar and only adding words within brackets where needed for clarity.

An in-person attendee, EO_2, who was also an event organizer, compared his HyWay experience to that with a social VR platform, *AltspaceVR*. Specifically, he found the 2-D map experience in HyWay (from a previous occasion when he was a remote participant on HyWay) to be better than the 3-D map experience in AltspaceVR. He stated, *"In a 3-D space like AltspaceVR, it is difficult to look around as I needed to frequently flip my avatar's view to know who is around me."* However, on a 2-D map he could see all the face bubbles and know who is around him in an easy and convenient way.

(b) Spatial awareness from porosity and face bubbles. Porosity provided an augmented spatial awareness to the remote participants. It provided a sense of being present in the crowd. In fact, remote attendee VA_3 stated, *"... the voice got louder as I got there [specific poster booth]. And I'm like, 'OK, I'm on the right way', and I got into the meeting, midway."* However, remote demoers frequently dialed porosity down to have a focused interaction. In this regard, remote demoer RD_2 stated, *"I kind of got overwhelmed with the other background voices, so I played with it [porosity slider] and made the porosity setting as little as possible ... but when I did it, it [background voices] just seemed like white noise. You know, like I could not [really] even make out words."*

Users found the face bubbles on the map beneficial in two ways. At a basic level, these provide an indication of the congregation of people in various locations, thereby enabling users to make navigation decisions accordingly. For instance, remote attendee VA_1 stated, *"You could see how many people were there in each room, which was nice because maybe you don't want to jump into a room and have a one-on-one conversation, but then there were other times where I was like, 'oh, that's great, this guys is like not so busy right now.' So, I'm going to go back and have a one-on-one conversation."* Furthermore, some users also found face bubbles beneficial as they were able to see *who* was where. For instance, remote demoer RD_1 stated, *"During the event, I was able to recognize people [from my organization] who were coming towards my booth. They too were able to recognize me and came to my booth to have informal conversation."* However, many participants mentioned that due to the nature of the event, they did not pay attention to the face bubbles as they were so many of them. One remote participant VA_1 said, *"I guess I viewed it [the event] more as like a 'star' interaction ... not so much across like attendee to attendee ... maybe if there was a coffee break room then I could have paid more attention to it [face bubbles]."*

(c) Hybrid vs. in-person experience. As mentioned before, all the participants stated that attending a hybrid poster event was a first-time experience for them. Importantly, all the remote participants liked the HyWay experience better than an in-person event. Remote attendee VA_2 said, *"I went [around] and talked to all of them [presenters] in a very short period-of-time without having to jump between multiple engagements."* Similarly, a remote demoer RD_1 stated, *"It [HyWay] gave me a booth like experience, one which you see during an actual [in-person] poster presentation."*

To probe the reasons for their overall positive experience on HyWay, we asked the remote participants to compare their experience on HyWay with that as an in-person attendee in a past poster event. VA_1 said, *"The last time I went to a tech conference was a couple years ago. And I remember one of the presenters had, like so many people around ... You're like looking through [peoples'] heads to try to see the [poster] board, and they're not quite big enough."* However, on HyWay he was able to view all the posters without any hindrance. He further stated, *"I really like the fact that I could see everything that I wanted without humans interfering my vision."*

One poster demoer, RD_2, called out the greater level of audio clarity he had compared to being an in-person presenter when interacting with an in-person attendee. He compared his experience with attending an in-person conference, where the noise level in a poster session overwhelmed his ability to interact with the audience. In this regard he mentioned, *"I found that at larger conferences it is difficult to have a conversation when there's lots of others around, louder conversations going around. So, this [HyWay] really hit the mark on that ... like just having a very clear [auditory] conversation."*

(d) Participants' engagement with the hybrid audience. All the demoers were able to engage with a variety of participants. Some participants like RD_1, and PD_3 mentioned that they were able to tag-team with their remotely present co-authors to engage effectively with the HyWay audience. Physical demoer PD_3 also mentioned that the map interface provided awareness about the remote participants. PD_3 stated, *"I was trying to be very aware of when people were appearing on that [map] screen."* To interact with them he leveraged one of his co-authors who was remote and spoke to the remote attendees. In this regard, PD_3 said, *"... I had a co-author who was remote ... there were times when I stepped further away from the system [poster screen] to take some questions from in-person attendees. My remote co-author spoke specifically to those who were attending virtually, and then we both had times where we covered each other on both fronts."* When asked about what if the in-person audience had some questions that only his co-author could answer, PD_3 mentioned that he was able to coordinate with his co-author when such a need arose. He said, *"I would say some of that coordination happened when someone [in-person] asked me a particular question. A question that my coauthor was in a better position to answer. I'd step closer to the screen and raised my voice slightly and then introduce the person [in-person attendee] and bring the two spheres together."* PD_3, however, emphasized that there is a learning curve in building such a coordination with remote users. He said, *"... [but] all this requires awareness of that envelope [hybrid space]. Like how far is too far, and how far is far enough."*

Though demoers like PD_3 were able to improvise their interaction with remote and in-person audiences, there were other demoers (both remote and in-person) who did face inconvenience when interacting with a mix of audiences. For example, PD_2 stated, *"On HyWay you have some of your audiences behind you [in-person audience] and some of your audiences in front of you [on the screen], so you're constantly whipping your head back and forth. Anytime, like one group, is always gonna feel left out. I had an in-person audience that kind of just walked by because I was presenting to someone online."* However, she further emphasized that such situations are also common during in-person poster events.

Some participants in the interview study also reported limited interaction between in-person and remote audiences. PD_1 attributed this limitation to the small size of remote users' video tiles. She said, *"I think it's more challenging for a remote person because they're very small on my screen, I think there's not a good way for them to communicate with anyone else except for me."* On a similar note, RD_1 (while visiting different poster zones as a remote attendee) felt that the in-person audience was more inclined to interacting with the physical presenters. The main bottleneck that he faced was his inability to reach out to the in-person audience. He stated, *"Who would initiate the discussion [at times] became a bottleneck on [HyWay]."* In this regard, PD_2 suggested that mannequins could be used to represent remote users. She said, *"Don't know if you need to just put like a mannequin and have that mannequin represent arrival of a remote person. That will help me understand who are the virtual presenters."* Based on this feedback, we believe there is a need to increase the salience of the remote participants in HyWay.

(e) System issues and effectiveness of off-the-shelf audio/video equipment. We sporadically faced system issues related to audio/video devices (e.g., lag from the browser app in launching camera and mic devices) throughout the event. This resulted in some of the remote participants exiting HyWay early. In this regard, virtual attendee VA_3, who faced such an issue said, *"At first, it was [uh] problematic because when I joined, for some reason I didn't have audio. I mean, I had audio. I could hear everybody, but nobody could hear me. I had to unjoin or leave and then come back and then my audio worked. Overall, it was a little disorienting at first because I wasn't sure since it was my first-time experience. I wasn't sure how I was supposed to behave."* Similar opinion were shared by the physical demoers. PD_1 mentioned that she lost some of the audience due to an audio glitch that happened at the start of the event. She said, *"... in the beginning because of those device issue I lost a few audience because the one who joined just couldn't hear or they couldn't see the screen, so they were completely lost."*

Interruption issues arising from off-the-shelf audio/video equipment also accounted for a majority of the poor experiences reported by the participants. The HyWay system leveraged built-in TV speakers and webcam-based mics [16]. EO_2, who was one of the event organizer mentioned that there is a small learning curve before we can provide a suitable audio experience to the participants. He said, *"There is some learning curve here ... as we get a bit more familiar with the system, we could personalize the audio experience for different demoers [remote/in-person]."* He highlighted some of the orientation issues in-person demoers faced with respect to the position of the webcam-based mic. EO_2 said, *"We did not have exclusive mics, instead we capitalized on the mics that were inbuilt inside the webcam we used ... I asked the presenter to face the camera as much as possible to avoid any voice break."* EO_1 and EO_2 both mentioned that in the future they plan to use parabolic mics [32] to enable a focused conversation between a remote and an in-person user. Though the event space had poster booths set ∼ 7 ft apart, both the event organizers mentioned that the use of webcam-based mics resulted in noise from the physical space that interfered with remote demoers' ability to have a focused conversation in their poster zones. EO_2 said, *"Webcam mics picked up a lot of noise from the event space. Unfortunately, those mics don't have inbuilt noise suppression."* Based on this feedback, we believe that there is an opportunity to use AI-based noise filtering and tune the mic sensitivity to improve the audience experience in the future.

9 DISCUSSION

9.1 Summary of Findings

First, the HyWay system design made it easy for the participants to use the system in all the five deployments. We found that the face bubbles on the map beneficial in two ways. At a basic level, these provide an indication of the congregation of people in various locations, thereby enabling other users to make navigation decisions accordingly. Furthermore, some users also found it beneficial to be able to see *who* was where. Also, all the participants found the 2-D map-based interface beneficial as it was informative and easy to navigate. For remote audience, porosity provided a sense of the event buzz and energy. Participants navigated the floor map and learnt about the different demo sessions either from the demo titles or by overhearing nearby presentations through porosity. However, a few remote participants reported dialing down the porosity level to have a more focused discussion during a poster session.

Second, from the interview study, we learnt that attending a hybrid event was a first-time experience for most of the participants. Nevertheless, most of the remote participants liked the HyWay experience better than an in-person event. They attributed this preference to a greater level of audio clarity, and the absence of obstruction by other users when viewing the poster titles and content. Importantly, both the in-person and the remote demoers were able to engage effectively with a variety of participants, both in-person and remote. In fact, for some demoers, the design of HyWay made them explore new ways to manage various participants. For example, one in-person demoer was able to tag-team with his remotely present co-author to engage effectively with the HyWay audience, both in the physical and virtual spaces. Though the overall experience of the participants was positive, some participants faced inconvenience when interacting with a mix of audiences. These participants mentioned that there is a learning curve involved in addressing both the in-person and remote audience concurrently as any one of them could feel left out when interacting with the other set.

Lastly, while the use of off-the-shelf audio/video devices was effective, this also accounted for most of the poor experience reported by the participants. Responses from the event organizers highlight the fact that providing personalized experience and quick troubleshooting of audio issues for different audiences requires greater familiarity than they had with the audio/video equipment as well as the space where the event is being organized.

9.2 Some Limitations of HyWay and Directions for Future Work

9.2.1 Localization of Physical Participants on the Virtual Map. Currently, virtual user avatars are tracked and depicted on the map continually, so other users are aware of their location and presence at all times. However, the physical participants, are only localized and their face bubbles depicted on the map when they are in a HyWay zone; their presence and movement outside of the zones is not tracked or depicted. In the future, we plan to explore ideas such as WiFi-based localization to accurately track the physical participants throughout the space and depict them on the map. We believe that the resulting enhanced awareness would promote greater interaction. For instance, once their presence and location is made readily and continually through physical user avatars, virtual users can navigate accordingly to connect with the physical users. Although HyWay would, at present, only allow them to converse in the designated physical zones.

9.2.2 *Directional cues.*

Awareness and navigation are central to the user experience in HyWay. While audio porosity aids awareness, at present it does not give any cues on the direction of the overheard audio. In the future, we plan to explore spatial audio, along with visual cues, to help users know where the source of the overheard audio is and thereby navigate towards it if interested. We believe this would approximate the real world, where a user's senses help them locate sound sources, thereby providing a more immersive experience to the remote users.

9.2.3 *Accessibility.*

While HyWay's design, including the map-based interface and audio porosity, work well for users who are sighted and able to hear, these could be challenging for users with disabilities such as the blind and the deaf. In the future, we plan to explore ideas such as tactile feedback, visual cues, and word clouds to make HyWay accessible to a wider cross-section of users.

10 CONCLUSION

People working remotely is no longer an exception and instead the new norm. At the same time, users value the *physical* workplace, conference, and other such settings. As we transition to this new hybrid world, we believe that enabling serendipitous and ephemeral conversations and mingling across physical and virtual users is important and a key shortcoming of existing conferencing systems.

In this paper, we have presented HyWay with the goal of addressing this need while making the experience equitable for all users — physical and virtual. The design of HyWay focuses on bridging the awareness gap, with techniques such as reciprocity, porosity, and map-based visualization and navigation, to enable a surprise-free interaction among users, just as when everyone is in-person. Furthermore, physical and virtual users in HyWay have the agency to move between conversations, just as users do in a physical hallway. The findings from the multiple deployments of HyWay are encouraging, with many of the features being well-received by users, while also pointing to shortcomings and the resulting opportunities for improving the system in future work. Among these are using spatial audio and visual cues to augment the porosity experience, and leveraging the computing substrate provided by HyWay to make hallway mingling and conversations accessible to the blind and the deaf.

ACKNOWLEDGMENTS

We thank the anonymous reviewers for their constructive comments, which greatly helped improve the paper. We would like to thank the Azure Communication Services (ACS) team — Alex Kwan, Artur Kania, Micah Vivion, Marius Ungureanu, Nikolay Muravlyannikov — and their leadership, Jim Kleewein and Bob Serr, for their support, especially as we exercised their platform in unconventional ways. We very much appreciate all the event organizers, Alex Blanton, Paul Bennett, Grace Higgins, Liz Zuidema, Johannes Gehrke, Sriram Rajamani and Kate Novak for inviting us to deploy HyWay in their events. We would also like to extend our gratitude to Johnny Johnson, Jordan Dahl, Gregory Lee, Jamie Lee, Rashmi KY, Vyshak Jain, Venkata Siva, Devaraj GR, Trevor Eberl and Kendall Martin for their active participation and help in deploying HyWay at various events in multiple global locations. We thank Sean Rintel, Lev Tankelevitch, Payod Panda and Bo Kang for their valuable inputs that helped shape the survey questionnaire and user studies. Last but not the least, we thank Rahul Bothra, Asim Baig, and Abhishek Dubey for helping us test the system on several occasions, and our study participants for providing us with a wealth of feedback and anecdotes.

REFERENCES

[1] 2016. *Sanbot*. http://en.sanbot.com/. [Online; accessed 27-April-2023].
[2] 2022. *Adobe Photoshop*. https://www.adobe.com/adobe/photoshop. [Online; accessed 27-April-2023].

[3] 2022. *Azure Communication Services*. https://azure.microsoft.com/en-us/services/communication-services/. [Online; accessed 27-April-2023].
[4] 2022. *Bezier Curve*. https://web.mit.edu/hyperbook/Patrikalakis-Maekawa-Cho/node12.html. [Online; accessed 27-April-2023].
[5] 2022. *Buying Guide: Comparing Field of View When Buying a Conference Room Video Camera*. https://www.videoconferencegear.com/blog/buying-guide-comparing-field-of-view-when-buying-a-conference-room-video-camera/. [Online; accessed 27-April-2023].
[6] 2022. *Creative SBS 240*. https://www.amazon.com/CreativeLabs-Creative-SBS-240-51MF0025AA003/dp/B0006SIL98/ref=cm_cr_arp_d_product_top?ie=UTF8. [Online; accessed 27-April-2023].
[7] 2022. *Discord*. https://discord.com/. [Online; accessed 27-April-2023].
[8] 2022. *Double 3, the future of Hybrid*. https://www.doublerobotics.com/. [Online; accessed 27-April-2023].
[9] 2022. *ExtraTreesClassifier*. https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.ExtraTreesClassifier.html. [Online; accessed 27-April-2023].
[10] 2022. *Face - REST API, Azure Cognitive Services*. https://docs.microsoft.com/en-us/rest/api/faceapi/face. [Online; accessed 27-April-2023].
[11] 2022. *FastAPI*. https://fastapi.tiangolo.com/. [Online; accessed 27-April-2023].
[12] 2022. *Fluent UI*. https://tinyurl.com/mtdrxbxh. [Online; accessed 27-April-2023].
[13] 2022. *Gather*. https://www.gather.town/. [Online; accessed 27-April-2023].
[14] 2022. *Inverse Square Law*. https://www.acoustical.co.uk/distance-attenuation/how-sound-reduces-with-distance-from-a-point-source/. [Online; accessed 27-April-2023].
[15] 2022. *LED screen viewing angles*. https://visualled.com/en/glossary/led-screen-viewing-angles/. [Online; accessed 27-April-2023].
[16] 2022. *Logitech Brio*. https://www.logitech.com/en-in/products/webcams/brio-4k-hdr-webcam.960-001105.html. [Online; accessed 27-April-2023].
[17] 2022. *Lowest Hearing Level*. https://cs.mcgill.ca/~rwest/wikispeedia/wpcd/wp/s/Sound.htm. [Online; accessed 27-April-2023].
[18] 2022. *Measuring Sound with Microphone*. https://www.ni.com/en-in/innovations/white-papers/13/measuring-sound-with-microphones.html. [Online; accessed 27-April-2023].
[19] 2022. *Microphone Specification*. https://invensense.tdk.com/wp-content/uploads/2015/02/AN-1112-v1.1.pdf. [Online; accessed 27-April-2023].
[20] 2022. *Phaser, Desktop and Mobile HTML5 game framework*. https://phaser.io/. [Online; accessed 27-April-2023].
[21] 2022. *Remo*. https://remo.co/. [Online; accessed 27-April-2023].
[22] 2022. *Sennheiser SK-100 G3*. https://en-in.sennheiser.com/sk-100-g3. [Online; accessed 27-April-2023].
[23] 2022. *SpatialChat*. https://www.spatial.chat/. [Online; accessed 27-April-2023].
[24] 2022. *Staying Visible When Your Team Is in the Office…But You're WFH*. https://hbr.org/2021/07/staying-visible-when-your-team-is-in-the-office-but-youre-wfh. [Online; accessed 27-April-2023].
[25] 2022. *Teams*. https://www.microsoft.com/en-in/microsoft-teams/group-chat-software. [Online; accessed 27-April-2023].
[26] 2022. *The Next Great Disruption Is Hybrid Work—Are We Ready?* https://www.microsoft.com/en-us/worklab/work-trend-index/hybrid-work. [Online; accessed 27-April-2023].
[27] 2022. *Tiled. Flexible level editor*. https://www.mapeditor.org/. [Online; accessed 27-April-2023].
[28] 2022. *Unhangout*. https://unhangout.media.mit.edu/. [Online; accessed 27-April-2023].
[29] 2022. *Visio*. https://www.microsoft.com/en-in/microsoft-365/visio/flowchart-software. [Online; accessed 27-April-2023].
[30] 2022. *Zoom*. https://zoom.us/. [Online; accessed 27-April-2023].
[31] 2023. *AltspaceVR*. https://altvr.com/. [Online; accessed 27-April-2023].
[32] 2023. *Parabolic Microphone*. https://en.wikipedia.org/wiki/Parabolic_microphone. [Online; accessed 27-April-2023].
[33] James R. Austin. 2018. Take a Researcher to Lunch: Informal Mentoring for Researchers. *Journal of Music Teacher Education* 28, 1 (2018), 6–9. https://doi.org/10.1177/1057083718802660
[34] Maral Babapour, MariAnne Karlsson, and Anna-Lisa Osvalder. 2018. Appropriation of an Activity-based Flexible Office in daily work. *Nordic Journal of Working Life Studies* 8, S3 (Apr. 2018). https://doi.org/10.18291/njwls.v8iS3.105277
[35] Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen. 2009. Pearson correlation coefficient. In *Noise reduction in speech processing*. Springer, 1–4.
[36] Sara A. Bly, Steve R. Harrison, and Susan Irwin. 1993. Media spaces: bringing people together in a video, audio, and computing environment. *Commun. ACM* 36 (1993), 28–47.
[37] Andriana Boudouraki, Stuart Reeves, Joel E Fischer, and Sean Rintel. 2022. Mediated Visits: Longitudinal Domestic Dwelling with Mobile Robotic Telepresence *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 251, 16 pages. https://doi.org/10.1145/3491102.3517640
[38] Valter Costa, Armando Sousa, and Ana Reis. 2018. Image-Based Object Spoofing Detection. In *Combinatorial Image Analysis*, Reneta P. Barneva, Valentin E. Brimkov, and João Manuel R.S. Tavares (Eds.). Springer International Publishing, Cham, 189–201.
[39] Paul Dourish and Sara Bly. 1992. Portholes: Supporting Awareness in a Distributed Work Group. 541–547. https://doi.org/10.1145/142750.142982

[40] Kevin Dullemond, Ben van Gameren, and Rini Solingen. 2011. Overhearing Conversations in Global Software Engineering - Requirements and an Implementation. *Food Chemistry - FOOD CHEM*, 1–8. https://doi.org/10.4108/icst.collaboratecom.2011.247099

[41] Kevin Dullemond, Ben van Gameren, and Rini van Solingen. 2010. Virtual Open Conversation Spaces: Towards Improved Awareness in a GSE Setting. In *2010 5th IEEE International Conference on Global Software Engineering*. 247–256. https://doi.org/10.1109/ICGSE.2010.36

[42] Robert Fish, Robert Kraut, Robert Root, and Ronald Rice. 1993. Video as a Technology for Informal Communication. *Commun. ACM* 36 (01 1993), 48–61. https://doi.org/10.1145/151233.151237

[43] William Gaver. 1991. Sound Support For Collaboration. https://doi.org/10.1007/978-94-011-3506-1_22

[44] S. M. Goza, R. O. Ambrose, M. A. Diftler, and I. M. Spain. 2004. Telepresence Control of the NASA/DARPA Robonaut on a Mobility Platform *(CHI '04)*. Association for Computing Machinery, New York, NY, USA, 623–629. https://doi.org/10.1145/985692.985771

[45] M.S. Granovetter. 1973. The Strength of Weak Ties. *The American Journal of Sociology* 78, 6 (1973), 1360–1380.

[46] Jens Emil Grønbæk, 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. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376379

[47] Jens Emil Grønbæk, Banu Saatçi, Carla F. Griggio, and Clemens Nylandsted 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

[48] Carl Gutwin and Saul Greenberg. 1998. Design for Individuals, Design for Groups: Tradeoffs between Power and Workspace Awareness. In *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work* (Seattle, Washington, USA) *(CSCW '98)*. Association for Computing Machinery, New York, NY, USA, 207–216. https://doi.org/10.1145/289444.289495

[49] Morten T. Hansen. 1999. The Search-Transfer Problem: The Role of Weak Ties in Sharing Knowledge across Organization Subunits. *Administrative Science Quarterly* 44, 1 (1999), 82–111. https://doi.org/10.2307/2667032

[50] Ellen Isaacs, John Tang, and Trevor Morris. 1996. Piazza: A Desktop Environment Supporting Impromptu and Planned Interactions. 315–324. https://doi.org/10.1145/240080.240316

[51] Bumsoo Kang, Inseok Hwang, Jinho Lee, Seungchul Lee, Taegyeong Lee, Youngjae Chang, and Min Kyung Lee. 2018. My Being to Your Place, Your Being to My Place: Co-Present Robotic Avatars Create Illusion of Living Together. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services* (Munich, Germany) *(MobiSys '18)*. Association for Computing Machinery, New York, NY, USA, 54–67. https://doi.org/10.1145/3210240.3210348

[52] Vahid Kazemi and Josephine Sullivan. 2014. One millisecond face alignment with an ensemble of regression trees. *2014 IEEE Conference on Computer Vision and Pattern Recognition* (2014), 1867–1874.

[53] Nuri Kim, Jeonghye Han, and Wendy Ju. 2014. Is a Robot Better than Video for Initiating Remote Social Connections among Children?. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction* (Bielefeld, Germany) *(HRI '14)*. Association for Computing Machinery, New York, NY, USA, 208–209. https://doi.org/10.1145/2559636.2563692

[54] Christoph Kirsch, Torben Wendt, Steven Van de par, Hongmei Hu, and Stephan D. Ewert. 2023. computationally-efficient simulation of late reverberation for inhomogeneous boundary conditions and coupled rooms. *journal of the audio engineering society* 71, 4 (april 2023), 186–201. https://doi.org/10.17743/jaes.2022.0053

[55] Robert Kraut, Carmen Egido, and Jolene Galegher. 1988. Patterns of Contact and Communication in Scientific Research Collaboration. *Proceedings of the 1988 ACM Conference on Computer-Supported Cooperative Work*. https://doi.org/10.1145/62266.62267

[56] Robert E. Kraut, Robert S. Fish, Robert W. Root, Barbara L. Chalfonte, I S. Oskamp, and Shirlynn Spacapan. 1990. Informal Communication in Organizations: Form, Function, and Technology.

[57] Robert E. Kraut and Lynn A. Streeter. 1995. Coordination in Software Development. *Commun. ACM* 38, 3 (mar 1995), 69–81. https://doi.org/10.1145/203330.203345

[58] Iris Lin and Catherine Kwantes. 2015. Potential Job Facilitation Benefits of "Water Cooler" Conversations: The Importance of Social Interactions in the Workplace. *The Journal of psychology* 149 (04 2015), 239–62. https://doi.org/10.1080/00223980.2013.874322

[59] Matthew Lombard, Frank Biocca, Jonathan Freeman, Wijnand IJsselsteijn, and Rachel J. Schaevitz. 2015. *Immersed in Media: Telepresence Theory, Measurement & Technology*.

[60] Debbi Long, Rick Iedema, and Bonsan Lee. 2007. *Corridor Conversations: Clinical Communication in Casual Spaces*. 182–200. https://doi.org/10.1057/9780230595477_9

[61] Cora Bagley Marrett, Jerald Hage, and Michael Aiken. 1975. Communication and Satisfaction in Organizations. *Human Relations* 28, 7 (1975), 611–626. https://doi.org/10.1177/001872677502800702 arXiv:https://doi.org/10.1177/001872677502800702

[62] Gregor McEwan and Saul Greenberg. 2005. Supporting Social Worlds with the Community Bar *(GROUP '05)*. Association for Computing Machinery, New York, NY, USA, 21–30. https://doi.org/10.1145/1099203.1099207

[63] Anthony J Myles, Robert N Feudale, Yang Liu, Nathaniel A Woody, and Steven D Brown. 2004. An introduction to decision tree modeling. *Journal of Chemometrics: A Journal of the Chemometrics Society* 18, 6 (2004), 275–285.

[64] Thomas Neumayr, Hans-Christian Jetter, Mirjam Augstein, Judith Friedl, and Thomas Luger. 2018. Domino: A Descriptive Framework for Hybrid Collaboration and Coupling Styles in Partially Distributed Teams. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 128 (nov 2018), 24 pages. https://doi.org/10.1145/3274397

[65] Mike Noon and Rick Delbridge. 1993. News From Behind My Hand: Gossip in Organizations. *Organization Studies* 14, 1 (1993), 23–36. https://doi.org/10.1177/017084069301400103

[66] Irene Rae. 2013. Using Robot-Mediated Communication to Improve Remote Collaboration. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (Paris, France) *(CHI EA '13)*. Association for Computing Machinery, New York, NY, USA, 1953–1956. https://doi.org/10.1145/2468356.2468709

[67] Irene Rae, Gina Venolia, John C. Tang, and David Molnar. 2015. A Framework for Understanding and Designing Telepresence. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) *(CSCW '15)*. Association for Computing Machinery, New York, NY, USA, 1552–1566. https://doi.org/10.1145/2675133.2675141

[68] Mehrnaz Sabet, Mania Orand, and David W. McDonald. 2021. Designing Telepresence Drones to Support Synchronous, Mid-Air Remote Collaboration: An Exploratory Study *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 450, 17 pages. https://doi.org/10.1145/3411764.3445041

[69] Samiha Samrose, Ru Zhao, Jeffery White, Vivian Li, Luis Nova, Yichen Lu, Mohammad Rafayet Ali, and Mohammed Ehsan Hoque. 2018. CoCo: Collaboration Coach for Understanding Team Dynamics during Video Conferencing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 4, Article 160 (jan 2018), 24 pages. https://doi.org/10.1145/3161186

[70] Kyoung Wan Cathy Shin and Jeonghey Han. 2016. Children's perceptions of and interactions with a telepresence robot. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 521–522. https://doi.org/10.1109/HRI.2016.7451836

[71] Anselm Strauss and Juliet Corbin. 1994. Grounded theory methodology. *Handbook of qualitative research* (1994), 273–285.

[72] Jaime Teevan, Brent Hecht, Sonia Jaffe, Nancy Baym, Rachel Bergmann, Matt Brodsky, Bill Buxton, Jenna Butler, Adam Coleman, Mary Czerwinski, Brian Houck, Ginger Hudson, Shamsi Iqbal, Chandra Maddila, Kate Nowak, Emily Peloquin, Ricardo Reyna Fernandez, Sean Rintel, Abigail Sellen, Tiffany Smith, Margaret-Anne Storey, Siddharth Suri, Hana Wolf, and Longqi Yang. 2021. *The New Future of Work: Research from Microsoft into the Pandemic's Impact on Work Practices.* Technical Report MSR-TR-2021-1. Microsoft. https://www.microsoft.com/en-us/research/publication/the-new-future-of-work-research-from-microsoft-into-the-pandemics-impact-on-work-practices/

[73] Michael Twidale. 2005. Over the Shoulder Learning: Supporting Brief Informal Learning. *Computer Supported Cooperative Work* 14 (12 2005), 505–547. https://doi.org/10.1007/s10606-005-9007-7

[74] Steve Whittaker, David Frohlich, and Owen Daly-Jones. 1994. Informal workplace communication: What is it like and how might we support it? 131–137. https://doi.org/10.1145/191666.191726

[75] Wikipedia. 2022. CIELUV — Wikipedia, The Free Encyclopedia. http://en.wikipedia.org/w/index.php?title=CIELUV&oldid=1102878349. [Online; accessed 27-April-2023].

[76] Wikipedia. 2022. YCbCr — Wikipedia, The Free Encyclopedia. http://en.wikipedia.org/w/index.php?title=YCbCr&oldid=1094228727. [Online; accessed 27-April-2023].

[77] Zhiwei Zhang, Junjie Yan, Sifei Liu, Zhen Lei, Dong Yi, and Stan Li. 2012. A face antispoofing database with diverse attacks. *Proceedings - 2012 5th IAPR International Conference on Biometrics, ICB 2012*, 26–31. https://doi.org/10.1109/ICB.2012.6199754