Digital Proxemics: Designing Social and Collaborative Interaction in Virtual Environments

ANONYMOUS AUTHOR(S)

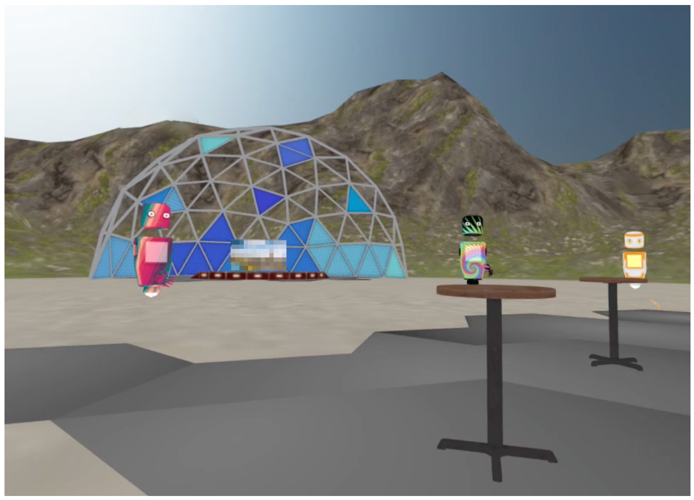


Fig. 1. Three avatars in Mozilla Hub's Outdoor Meetup Space.

Behaviour in virtual environments might be informed by our experiences in physical environments, but virtual environments are not constrained by the same physical, perceptual, or social cues. Instead of replicating properties of physical spaces, we can manipulate dynamic environmental, aural, and social properties in virtual experiences that diverge from reality. This paper lays a foundation for *digital proxemics*, which describes how we use space in virtual environments and how the presence of others influences our behaviours, interactions, and movements. First, we describe the open challenges of digital proxemics in terms of human activity, social signals, audio design, and environment. We explore a subset of these challenges through an evaluation that compares two audio designs and two displays with different social affordances (head-mounted display versus desktop PC). We use quantitative methods using instrumented tracking to analyse behaviour, demonstrating how personal space, proximity, and attention compare between desktop PC and head-mounted display.

Additional Key Words and Phrases: Virtual Environments, Digital Proxemics, Social Signal Processing, Quantitative Methods.

ACM Reference Format:

Anonymous Author(s). 2022. Digital Proxemics: Designing Social and Collaborative Interaction in Virtual Environments. In *CHI '22: ACM Symposium on Computer Human Interaction, May 01–05, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Digital proxemics describe how we use space in virtual environments (VEs) and how the presence of others influences our behaviours, interactions, and movements. Proxemics in physical environments has been extensively researched [15, 19, 23]. Research on the relationship between proxemics and technology is also well established [17, 28]. *Digital proxemics* concerns a distinct and emerging area of research [5, 26, 44] that concerns itself with how social proximity is perceived and acted upon in malleable virtual environments. Recent advances in the availability, fidelity, and immersive displays for experiencing virtual environments have expanded the possibilities for research in this area.

Understanding how we use space in VEs builds upon but is distinct from proxemics in physical settings. VEs do not have the constraints of physical environments and can be manipulated and reconfigured in real time. For example, the size and layout of virtual rooms can be altered during interaction, changing the perceived crowdedness or cosiness of a virtual space. Audio parameters can be changed to amplify a speaker and minimise noise from the audience. Engineering successful virtual experiences will depend on being able to make such changes with intention and design, where poor decisions will result in unfit or unusable spaces. Additionally, the range of affordances and social cues available in virtual environments is dramatically different from physical environments. We experience a VE through an avatar, which may have less articulation than our own bodies. The field of view may be narrower than human vision, giving a lower bandwidth visual channel to perceive our surroundings. Senses that play a key role in physical proxemics, such as body odour and skin warmth, are often completely missing from current VEs. The range of social signals possible in a VE may be limited, but control over how we present ourselves in a VE is much more flexible than in physical settings. For example, pre-rendered animations performed by avatars allow us to give off desired signals without physically performing them. Physical proxemics provide the inspiration, but the beyond reality capabilities of virtual environments present a new challenge for designing virtual experiences.

In this paper, we lay a foundation for *digital proxemics*, discussion the open challenges in this emerging area in terms of human activity, social signals, audio design, and environment design. Understanding the patterns of behaviour for the activities we intend for virtual environments provides a key starting point [15, 23]. Do we want to facilitate serendipitous networking events [36] or focused small group discussions [2]? These patterns of behaviour may unfold differently in virtual environments when we consider the *social signals* and non-verbal behaviours available. *Audio design* represents a key modality, second only to vision, for interaction and perception in virtual environments. And the environment itself, which provides the backdrop for the entire experience, will influence how interaction unfolds.

To begin exploring *digital proxemics*, we conducted an evaluation in a virtual environment using Mozilla Hubs[1], an open source VE platform that runs in a standard browser. Our evaluation compared two audio designs and two display types supporting different affordances for social signals. The first audio design, called "cocktail," simulates the physical audio environment of a cocktail party where background voices are audible throughout the VE. The second audio design, called "bubble," silences background talk beyond eight meters. We also compared desktop PC display to head-mounted display (HMD), where HMD users had greater expressivity through head movements and tracked hands compared to desktop PC keyboard controls.

Addressing research in *digital proxemics*, we demonstrate the following contributions: (i) a proposed research area for*digital proxemics* organised into key factors of activity, social signals, audio, and environment, (ii) a comparison of two audio conditions, measuring the impact of background noise on group proximity, and (iii) a comparison of

[1] Mozilla Hubs: https://hubs.mozilla.com

behaviour using an HMD versus a desktop PC, quantifying different behaviours in terms of personal space and social signals demonstrating attention.

2 RELATED WORK

2.1 Proxemics in Physical Spaces

This research builds on Hall's foundational work on proxemics [19], which describes human proxemics in face-to-face interactions. Proxemics in physical spaces is based in how people perceive distance using their eyes, ears, skin, and even noses [19] to determine a comfortable physical distances. How much of another person's body is visible in your field of view, how loud you perceive their voice to be, whether you can feel the heat of their body, or even smell their cologne, all factor into negotiating personal space during face-to-face interactions. This sense of personal space is dynamic, and may change depending on context, culture, and behaviour.

Hall defines four proxemic zones; intimate, personal, social, and public [19]. In the *intimate* zone (< 0.46 meters), physical contact may heighten or distort social cues, field of view will be close to the other person's face, you might feel their breath or the heat of their body. In the *personal* zone (0.46–1.2 meters), you can still reach out and touch someone but physical contact is not constant. This relative closeness makes facial expressions, movements of the eyes, and other small movements more pronounced. In the *social* zone (1.2–3.6 meters) you would not expect any physical contact and can view the other person's whole body more fully as they move further through this zone. In the *public* zone (> 3.6 meters) there is decreasing visibility of the face and audibility of the voice but more of the periphery opens up.

The proxemics of face-to-face interactions have been analysed in terms of attributes beyond physical distance. Kendon describes *F-formations*, introducing the relative orientation of a group as an important attribute of proximity. F-formations arise during encounters that are sustained between two or more people in close proximity where they are oriented towards a shared space with exclusive, direct, and equal access [23]. Goffman's approach has less emphasis on physical spacing and instead analyses the way people allocate and demonstrate their attention and focus [16]. Goffman's research describes face-to-face interactions as *unfocused* or *focused*, with special consideration to focused interactions where groups gather together to collaborate on a single goal or point of attention. Focused interactions can be complex, with tight or loose social regulations, explicit and implicit boundaries for participation, and different expectations or affordances for involvement [16]. Whyte focuses on the design and affordance of urban spaces [42], considering how the availability and design of sitting space, relative position of the street, and exposure to sun and wind change where and how people gather and meet in public settings. Whyte's research on *effective capacity* and perceived *crowding* combine physical space design with human proxemics. Physical capacity can be very different from *effective capacity* when space is designed well, for example small spaces like Greenacre Park in New York can feel less crowded because the loud water feature masks the noise of other people. Whyte discusses how effective capacity is dynamic and self levelling based on proximity of others, comfort, and amenities.

Translating physical proxemics into VEs presents challenges because social and environmental cues may lower fidelity or completely absent when translated to the virtual. Olfaction and scent play a significant role in physical proxemics, but are challenging to incorporate in virtual environment. Haptics have similar limitations, with capabilities for tactile, thermal, and force feedback in virtual environments. Audio experienced through simulation and loudspeaker may be significantly different from audio in physical environments [39]. On the other hand, proxemic cues in virtual environments may be deliberately manipulated, distorted or enhanced to alter the social dynamics of a virtual space.

2.2 Interaction and Personal Space in Virtual Environments

In HCI, proxemics have been applied to the design and analysis of interactive systems [17] and to understand how technology plays a role in physical proxemics [28]. There is an established body of research exploring how social signals from the physical world translate into the virtual, and how this impacts user experiences. The range of affordances, particularly around embodiment and the social signals available in different platforms, is a key factor in how a VE can support collaboration and social interaction [9]. Benford et al. look holistically at embodiment in virtual environments [4], discussing a broad range of issues including how users can demonstrate their presence, position, orientation, facial expression, and identity. Many of the issues, such as assessing whether another person is *actually* present, persist in current VE applications. Guye-Vuillème et al. investigated non-verbal cues in a collaborative virtual environment, allowing participants to trigger postures, facial expressions and gestures [18]. Bowers et al. focused on talk and turn taking given limited affordances for social signals in a VE called MASSIVE [6], finding that participants quickly adapted to the abilities of the "blockies" avatars to anticipate turn-taking and negotiate interactions. Moore et al. consider how environment design impacts of social activity in virtual space, identifying accessibility, social density, activity resources, and hosts as key factors and creating successful virtual spaces [30].

Personal space in virtual environments is a key issue for digital proxemics, especially when platforms provide inconsistent mechanisms for establishing and protecting personal space. Hecht et al. completed comparisons on the shape of personal space in physical and virtual environments, finding that personal space was roughly circular and consistent between real world and virtual encounters [20]. Wilcox explored discomfort when personal space is violated in virtual environments, demonstrating significant negative reactions comparable to the same experience in physical environment [43]. Llobera et al. analysed physiological arousal using skin conductance when participants were approached by virtual characters at different proximities [26],demonstrating heightened physiological arousal the closer virtual characters approached. Bailenson et al.[3] explored how gaze impacted personal space between participants and virtual agents, finding that participants avoided collisions with the virtual agent. The gaze of the virtual agent also impacted their performance in a memory task in comparable ways to physical observers. Podkosova et al. explored proxemics and locomotion when participants were in a shared virtual environment where some participants where physically co-located and others were distributed [34] For distributed participants, collisions were more common and sense of co-presence lower than between co-located participants.

Proxemics have also been used to enforce personal space and protect users in a virtual environment. Pohl describes how proxemic zones could be used to enforce rules on which objects and agents are allowed in personal space [35]. McVeigh-Schultz et al. analyse the properties of a range of virtual environments in terms of embodiment, social mechanics, and functions for shaping and enforcing social norms to prevent harassment [29]. The ethics of "immoral behaviour" and misrepresentation remain open challenges in virtual environments [7].

2.3 Methods for Digital Ethnography in Virtual Environments

Evaluating interaction in virtual environments draws heavily from ethnography, with a significant body of work using observation and interview methods to analyse user experience. Ethnographic methods have been applied to virtual environments like World of Warcraft [32], There [8], Second Life [12], and multi-player dungeons [1]. These approaches often make use of *participant observation*, where researchers actively engage with the environment and other users to gather qualitative data on their experiences.

Virtual environments also afford automatic logging and collection of quantitative data. In commercial platforms like Second Life, researchers have used bots to log extensive data about interaction and movement through the virtual world [40]. For example, Varvello et al. found that avatars in Second Life formed small groups similar to physical settings using this quantitative approach [40]. Friedman et al. also used bots in Second Life [12], using proxemics to analyse behaviour when players were approached by logging bots and forming small groups. Fraser et al. completed a lab study to compare virtual and physical environments in terms of field of view, haptic feedback, and latency [11]. Schroeder et al. describe two quantitative methods using interaction logs and manually tags data for statistical analysis of interaction in a VE [37], focusing on categorising activity types and occurrences of events. Combining qualitative and quantitative methods using observation, logging, interviews, and surveys have also proven effective. Ahn et al. completed a survey to analyse user experience at an academic conference, focusing on social presence and satisfaction with the virtual conference format[2]. Williamson et al. completed quantitative logging with qualitative interviews to analyse interaction in a workshop event [44]. Le et al. gathered quantitative data through surveys and logs across multiple platforms and qualitative observation data to analyse comfort, motivation, and experience during an academic conference [24].

3 DIGITAL PROXEMICS

Understanding *digital proxemics* will change the way we design interaction for virtual environments, creating opportunities for richer and more varied social experiences. In virtual environments, we can manipulate interaction to facilitate social and collaborative activities beyond those possible in physical environments. To meaningfully engineer these VE experiences, we need to understand how people react to social cues in virtual environments, how changing virtual parameters impacts behaviour, and how to apply these to construct effective virtual social interactions. For example, what parameterisation promotes the most serendipitous interactions, or the most focused ones? Which configurations encourage equal participation, or alternatively give leaders more power? Understanding how to design and configure virtual environments to promote or constrain social behaviours is the core of *digital proxemics*. Taking inspiration from physical proxemics, we propose a research space for *digital proxemics* based on activity, social signals, audio, and environment.

3.1 Activity

The structure of social activities provides a starting point for analysing *digital proxemics*. This might include the number of focus points in a shared space, number of participants, and expectations around participation. Designers often aim to facilitate specific activities like presentations, networking events, or small group collaborations. The needs and constraints of a presentation from a single person to a large group are dramatically different than a networking event with no fixed focus point. Goffman describes interaction as *focused* or *un-focused* [16]. Kendon further studied *focused* interactions in substantial detail [23].

3.1.1 Unfocused Interactions.

Unfocused interactions happen where people are co-present but not engaged in a shared activity; for example, wandering about a networking event. Unfocused interaction is primarily non-verbal, but people may “give off” complex behavioural signals as part of managing co-presence. Their posture and facial expression could indicate availability for interaction. Their position and body orientation could communicate their intention to join an existing group. Well engineered unfocused interactions play a key role in familiarisation with a virtual environment, serendipitous encounters, and forming groups.

3.1.2 Focused Interactions. Focused interactions involve groups engaged in coordinated activities; for example conversing as a group or listening to a presentation. Focused interactions cover a broad range of social activities [23], which may or may not have highly formal rules around access, turn taking, attention, and participation [16]. Understanding the formal and informal rules of the intended activity and how social norms are negotiated can inspire VE designs. For example, a presentation with a single focus point may be improved by amplifying the presenter's voice and muting observers. In less structured focused interactions, for example breakout groups, more flexible affordances might be needed.

3.2 Social Signals

In all interactions, the non-verbal signals we continuously *give* or *give off* are a rich source of information that is constantly interpreted by others. These signals can be understood in terms of information theory [23], where some have greater bandwidth, some are more noisy, etc. The range of social signals available in a VE can be dramatically different than face-to-face interaction, and we are often missing signals (notably scent [19]) or are restricted to transmitting them in limited or awkward ways [5].

The design of the VE or the mode of interacting with the VE can impact users' performance and perception of social signals. Many VEs support interaction on both desktop PC and head-mounted displays. The additional expressiveness of hand tracking or ease of visual scanning with an HMD create very different experiences [44]. In physical spaces, Kendon discusses the ease of maintaining mutual eye contact, of positioning the body, and of gesturing, and how easy it is the observe these in turn [23]. There may not be one-to-one relationships with these signals when they are translated to the virtual. Is some ways VEs can enhance our perceptions, for example notifications when others enter and exit a room. Such signals may be more accurate and more effective than the physical equivalent of watching the door.

3.2.1 Body Position and Orientation. The basis of physical proxemics is relative body position. Refinements such as F-formations introduce relative orientation. These attributes can directly translate to a VE—avatars have a "front," they take up space, they can move. But there is also potential for beyond reality interaction, for example presence in multiple locations, instantaneous movement, and omni-directional sensing. Matching these "beyond reality" capabilities to human social dynamics represents an open design challenge for interaction in VEs.

3.2.2 Articulation of Head, Body, and Limbs. The range of non-verbal communications affects the experience of a VE [4], but how these are generated or performed does not need to be a 1:1 relationship with physical movement. For example, *Emotes* are user-triggered canned animations (such as waving, dancing or other gestures) to communicate non-verbally. Emotes require less effort and give individuals more control than physical actions.

3.2.3 Facial Expression. Hall describes the key role that facial expressions play in intimate and personal proxemic zones [19], but the effects of restriction on field of view in close proximity will be different in a VE. Facial expression is also commonly absent or weakly presented in VEs, although it is increasingly being explored [38].

3.2.4 Physical Appearance. The body-ownership and embodiment of avatars has been extensively researched [10, 27, 33]. The ability to customise appearance, and the range of options available, will also have an impact on *digital proxemics*. For example, previous work has found differences in how users maintain space between humanoid and abstract agents in virtual environments [26].

3.3 Speech and Audio Design

Audio is one of the main modalities (after visual) for engaging in a VE, with many possibilities for beyond-reality interactions. The production, perception, and interpretation of speech has been widely researched in physical and simulated environments. Thery et al. found that binaural audio presented over headphones and loudspeakers differed significantly in perceived reverberation and listener envelopment [39]. For example, these changes in sound perception could have significant impacts for how the "cocktail party effect" [45] is experienced in physical or virtual environments. Gil-Carvajal evaluated audio perception when presented with incongruent visual or auditory cues [14], finding the mismatches could significantly disrupt distance judgements.

Numerous approaches have been used for simulated and propagating sound in virtual environments [13], but how these impact behaviour and user experience is still an open challenge. The ability to dynamically manipulate audio parameters makes this an interesting part of digital proxemics, is straightforward to achieve in current VEs, and potentially disrupts our perception and social signals in ways that needs careful design. For example, a group needing privacy in a physical setting might huddle in a corner. In a VE, one could create a private zone which may or may not give off signals for privacy to others. Audio design could also create super powers, for example whispering across distances, creating instant megaphones, and breaking apart the relationship between distance and audibility in both useful and confusing ways.

3.4 Environment Design

How we design space is crucial to how we use space [21, 42]. Virtual environments diverge from physical environments in significant ways, creating infinite possibilities in terms of form, malleability, scale, and functionality. Virtual environments may also lack familiar aspects of physical environments, where collisions, boundaries, and other physical properties may be limited or completely absent. Moore et al. describe the challenges of designing successful virtual places in face of limitless possibilities [30], analysing how the accessibility, social density, activities, and hosts impact the success of virtual spaces. Moore et al, discuss the tendency to create complex spaces because *we can,* but complex or expansive spaces often make poor virtual places.

Virtual environments also afford beyond reality designs that could be leveraged for better interactions. For example, non-euclidean virtual environments [31] can be navigated in ways not possible in physical environments. Imagine a hallway to breakout rooms that is long to enter but short to return. Would increased travel time to breakout rooms aid in group formation and cohesion? Playing with virtual space is an open challenge in digital proxemics where our models of proximity are still grounded in physical environments.

4 DESIGN AND EXPERIMENT

As our first steps in exploring *digital proxemics*, we identified two factors which can affect how small group discussions in a virtual environment: audio attenuation model and display modality. We compared audio models where background noise is always present versus background noise silent beyond eight meters. We also compared desktop PC display versus head-mounted display. We selected these conditions based on three hypotheses:

H1 People using HMDs will make more use of personal space, collaborating at a further distance due to embodiment and hand gesturing.

H2 People using HMDs will use more social signals to express attention and maintain peripheral awareness through head movement.

Table 1. Participants were grouped by *audio condition* and *display modality*. The audio conditions (A) Cocktail (or Inverse) and (B) Bubble (or Exponential) and display modalities (I and II) and are detailed in § 4.2. Each group consisted of two parts (for example 1.1 and 1.2 represent both discussion groups of 3 people each in Group 1).

Audio Condition	(I) **Head-mounted Display**	(II) **Desktop PC**
(A) **Cocktail**	Group 1	Group 3
(B) **Bubble**	Group 2	Group 4

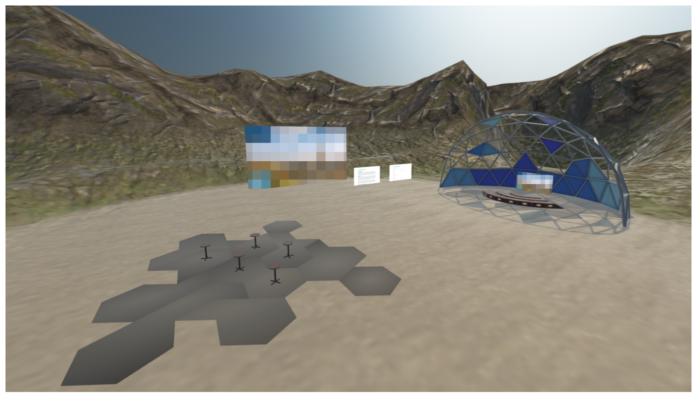


Fig. 2. All experimental sessions were held in the Mozilla Hubs "Outdoor Meetup" environment. The virtual space is a large open outdoor environment measuring seventy by forty meters with some small tables, a floor decal, and a small dome with amphitheatre seating.

H3 Environments with background noise will make small group discussion more challenging and distracting.

4.1 Virtual Environment

We used the open source Mozilla Hubs as our experiment's VE. This allowed us to easily modify the client code running in the HMD or desktop browser. This includes instrumenting avatar's positional data (x, y, z), forward vector, and various state flags (flying, muted, etc.) at each frame. We built our data collection platform for Mozilla Hubs by modifying a previously open sourced Hubs research collection framework [25]. All of our code, data, and analysis scripts are open sourced[2]. For the experiment, we used a single environment across all the sessions—an openly available virtual room in Hubs called "Outdoor Meetup." as seen in Figures 1 and 2. We added elements specific to our evaluation, such as links to the information sheet and questionnaire and worksheets. The Outdoor Meetup space is a large outdoor environment measuring 70×40 meters and has been used an several different evaluations and events [2, 24, 44].

4.2 Experimental Conditions

We ran a between-subjects evaluation with two factors: audio design and display modality. For audio design, we compared a distance attenuation model where background talk is audible (inverse model) or inaudible (exponential model). We also compared desktop PC display to head-mounted display to compare different affordances for social signals.

[2]Available upon publication to maintain anonymity.

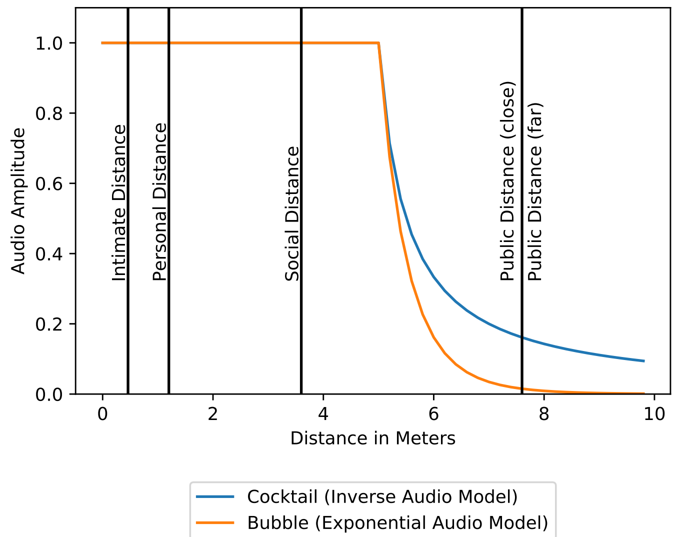


Fig. 3. Attenuation per meter in the two audio conditions (linear scale). Vertical lines indicate Hall's zones [19]. While the Cocktail/Inverse condition A asymptotically approaches x-axis after the public distance, the Bubble/Exponential B model descends faster after 5 meters and crosses the x-axis at 10 meters.

4.2.1 Audio Design. We explored if manipulating audio design affects social collaboration in virtual environments. Many features affect speech transmission and intelligibility [22], and we focused on the attenuation of background noise. We measured how people establish interpersonal distances in conversations in the presence of background sound. To control background sound levels, we adjusted Hub's attenuation model to follow either an inverse or an exponential model.

Audio Condition A. *The Cocktail Inverse Model attenuates audio similarly to the default Hubs configuration and the "ideal audio" simulation optimised to flatten at the "public distance" zone. The effect is similar to background chatter one would pick up at a cocktail party or other social gathering.*

Audio Condition B. *The Bubble Exponential Model has a faster decay and silences audio beyond a fixed distance threshold. This is similar to being inside an audio bubble.*

Condition (A) is close to the attenuation model in the physical world. Condition (B) creates an audio isolation field around each avatar, unlike that experienced in the physical world (however, an analogous effect can be seen in in museums with hyberbolic shielded speakers pointing down to make listening areas around media installations). Figure 3 shows the audio attenuation curves for the two conditions.

4.2.2 Social Signals and Display Modality. We explored if display modality affects proxemics. The display device constrains the social signals that can be communicated and perceived. We compared interaction with a desktop PC (controlled by mouse and keyboard) and a head-mounted display with hand held controllers.

Display Modality I. *This condition used an Oculus Quest HMD. Immersive experiences with HMDs allow avatars to express head movement with continuous tracking through the HMD. Avatars also have continuously tracked hands which* *will animate and move through the space as a user gestures and talks. This hand space may create an added personal perimeter to an avatar.*

Display Modality II. *This condition uses a conventional desktop PC. Movement of the avatar and head are controlled using keyboard and mouse. Interaction is discrete, where movements are only triggered while the users actively manipulates keyboard and mouse. The avatar does not have hands.*

4.3 Procedure

Each group consisted of six people with two authors facilitating the discussion. At the beginning of each session, participants were welcomed by the facilitators and familiarised with the virtual environments. As a large group, participants were presented with a consensus seeking task to complete in small groups.

We chose a consensus seeking task [41], where people can verbally engage with each other as single group or multiple sub-groups. This is in contrast to puzzle-based tasks which require interaction with an object or device. The consensus task was to agree upon a subset of items from a list. Participants were asked to imagine they are lost at sea with only a life raft, a book of matches, and their fellow group members. They must jointly select five items from a list of fifteen items that would maximise their chance of survival at sea. Table 2 details the full task and items the participants could select. The participants were asked to split into two groups of three to complete the task.

After completing the small group task (10–15 minutes), all participants again formed a single group with the facilitators to present their decisions and review the suggested best solution. The session was concluded with a group discussion about the experience overall and an exit survey.

4.4 Participants

We recruited 24 participants through mailing lists, social media, and local networks. Each prospective participant was given a short screening questionnaire and the selected participants were formed into four groups of eight participants (see Table 1) by display modality and audio condition. Each session lasted approximately 45 minutes and participants were monetarily compensated for their time. Participants were instructed to arrive ten minutes before session start time to complete technical checks, familiarise themselves with the room, and were advised to use a network with sufficient speed (e.g. avoid slower public WiFis). The experiment was reviewed and approved by an institutional ethics committee.

4.5 Results

In our customised instance of Mozilla Hubs, client-side logging is completed at the individual's frame rate and can vary depending on hardware speed and network condition. This results in a variable frame rate of logged data per participant which may change over time. To correct this, we resampled the time series from each participant to 30 frames per second (fps). After resampling, there were 3,034,125 events generated from 24 participants plus the 2 facilitators during 4 sessions.

4.5.1 Audio Design. The key difference in our two audio conditions was the presence of background noise and chatter when the participants split into two groups for the consensus seeking task. In the cocktail condition (A), participants could always hear the other group in the background no matter how far apart they stood. In the bubble condition, (B), participants would not hear the other group if they stood more than eight meters apart. Although background noise would be realistic in a physical environment, we hypothesised that the bubble without background noise would promote distance between avatars (and provide fewer distractions for small group focused interactions).

Table 2. The collaborative task worksheet given to the participants of each group.

Sea Survival Worksheet

Consider the following 15 items and their usefulness if you were lost at sea. As a group, select the five most important items you would choose to maximise your chances of survival. You must agree as a group on the final list you would select.

- Sextant (A navigation instrument for measuring angular distances)
- Shaving mirror
- Five-gallon can of water
- Mosquito netting
- Once case of army rations
- Maps of the Pacific Ocean
- Seat Cushion
- Two-gallon can of oil-gas mixture
- Small transistor radio
- Shark repellent
- Twenty square feet of opaque plastic
- One quart of 160-proof Puerto Rican rum
- Fifteen feet of nylon rope
- Two boxes of chocolate bars
- Fishing kit

Figure 4 details each group's distances during the small group discussion. Using pair-wise distance calculations for small group interaction segments, we analysed the standing distance maintained during these discussions. Examining the proxemic zones for cocktail versus bubble audio conditions, we see that for both HMD and desktop PC the cocktail mode pushed participants closer. Particularly for the desktop PC participants, the majority of interactions occurred within the personal zone and even collided in the intimate zone. HMD users were much more likely to space themselves within the social zone. This behaviour, almost like leaning in to better hear, resulted in tighter small groups in the cocktail audio mode.

4.5.2 Social Signals and Attention. When interacting with the VE using a head-mounted display, participants have more affordances for giving off social signals, and this was reflected in how they positioned themselves and demonstrated their attention during the small group discussions. Figure 5 visualises a top-down view of participants' perspectives in HMD versus desktop PC. Each participant is visualised from the centre-point, and other participants are shown as a scatter plot. When a participant microphone is activated, the plot is coloured red (grey otherwise). Figure 5 demonstrates how participants using an HMD moved their head to keep other participants in their field of view, especially when others were speaking. In contrast, desktop PC users maintained a more static field of view and did not always turn to face the active speaker.

4.5.3 Personal Space. To analyse personal space when different sizes of groups form, we analysed the closest standing person while interacting in small (3 participants) and large (6 participants, 2 facilitators) groups. Analysing just the

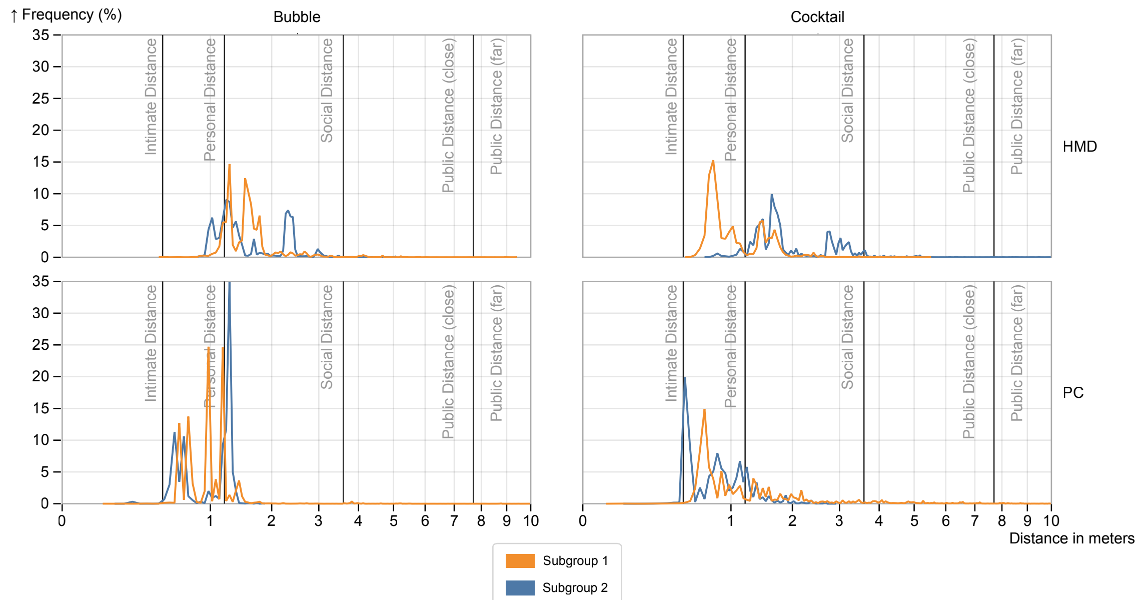
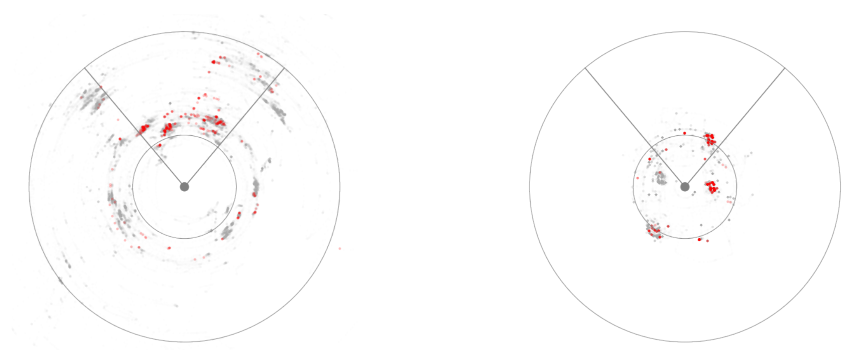


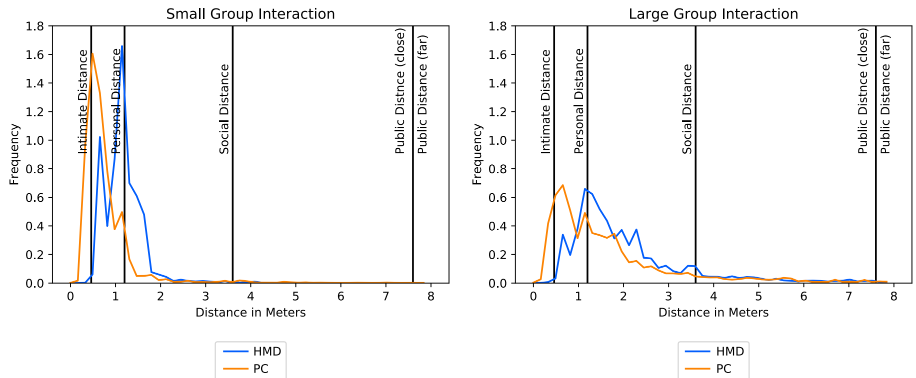
Fig. 4. Using pairwise distance calculations at 30 FPS during small group discussions, this visualisation shows the distances small groups stood within during the consensus seeking task. Desktop PC participants stood closer together, mostly occupying intimate to personal distances. By contrast, HMD participants maintained personal to social distances. The two line colours denote the sub-groups in each condition. The x-axis is on a logarithmic scale.



(a) Small Group HMD. Participants kept others in their field of view, especially when speaking.

(b) Small Group Desktop PC. Participants did not keep others in their field of view, sometimes even allowing others to stand behind them.

Fig. 5. Top-down view of each participant's perspective overlayed in a single scatter plot. When microphone is activated, points is plotted in red (grey otherwise). Inner circle visualises beginning of personal distance. Outer circle visualised beginning of social distance. Arc visualises field of view.



(a) Closest standing person during small groups.

(b) Closest standing person during large groups.

Fig. 6. Personal space as shown by closest standing person based on pair-wise comparison of standing distance during small and large group activities in a density plot. In large groups, participants claim more personal space, even extending into the public proxemic zone.

nearest person, rather than all others, allows us to measure personal space independent of group size (larger groups necessarily will take up larger spaces). The nearest person gives a metric for how much space each person is comfortable maintaining in any size group. Figure 6 compares the nearest person for large and small groups using HMD and desktop PC.

The small groups created closer formations, with desktop PC participants gathering at a more intimate distance compared to HMD users who kept a personal distance. Given the additional social signals available using HMD, we expected to see HMD participants making use of more personal space, with more sensitivity to collisions in the intimate zone. While using an HMD, participants have better articulation of the head and hands. Hall describes the distance of arms reach as crucial for delimiting the intimate zone [19], and the absence of hands in the desktop PC condition resulted in many more collisions in the intimate space.

When forming a large group, participants often claimed more personal space. In particular, HMD participants were much more likely to stand in the social proxemic zone while desktop PC participants were still likely to crowd in the personal zone. During the large group interactions, there were also notably interactions in the public zone. Even though these discussions were focused, the larger group was not held together as strongly as the small groups.

5 DISCUSSION

We evaluated social interactions in a virtual environment comparing desktop PC and head-mounted displays under two audio designs (one matching physical acoustic reality, the other breaking reality).

Comparing HMD to desktop PC was an important and obvious first step. We hypothesised that HMD participants would make greater use of personal space and benefit from the enhanced social signals (H1). Our results demonstrate that HMD users were more conscious of their personal space, avoiding collisions in the intimate and personal zones, orienting their bodies towards others, and keeping the active speaker in their field of view. These behaviours better reflect expectations for interaction in physical spaces [19, 23], compared to desktop PC participants. Participants in the HMD condition consistently maintained a larger personal space compared to desktop PC, supporting our first hypothesis.

When we analysed perspective and field-of-view (Figure 5, we found that HMDs participants were more likely to keep others in their field of view, especially when they were speaking. This supports our second hypothesis as participants kept speakers in focus and maintained avatar-to-avatar (read: eye-to-eye) contact (H2). This raises an open question around how much bringing in real-world behavioural metaphors improves user experience. For example, does performing social signals for attention make turn taking easier? In the absence of these social signals, would we observe more cross-talk and conflict? We did not record speech in this evaluation due to privacy concerns, but adding speech data streams to this approach would add a valuable dimension in the future.

While some of the proxemic differences we observed can be attributed to the hand-gesture space, it is likely there are also perspective effects in an HMD which lead users to feel uncomfortable when others are too close. While the field of view in the VE is rendered at the same angular width for HMD and desktop PC users, it is experienced very differently. One can always sit back from a monitor and put the VE in a smaller portion of their true field of view. This warrants further exploration. If desktop PC participants are required to maintain a fixed distance from the display (commonly achieved using head rests in perceptual studies), would we observe an increase in personal space? Our study was completed remotely with distributed participants as they would realistically interact with a VE, but more controlled lab experiments could address fundamental perceptual questions.

We hypothesised that background noise would make small group interactions more challenging (H3), but our results are more inconclusive on this hypothesis. We measured participants crowding closer in the "cocktail" mode in both HMD and desktop PC conditions, almost like leaning in to better hear. However, we do not have enough data to assess exactly how audio design impacted more qualitative aspects of experience. For example, how does audio design impact peripheral awareness or distraction? Does the cocktail party effect [45] work similarly in a virtual environment compared to physical settings? By instrumenting the virtual environment alone, we cannot make definitive claims on H3 yet. Anecdotal evidence[3] indicates that audio design and background noise have significant impacts on user experience, but more research is needed here.

One factor the remains untouched in this evaluation is the impact of environment on *digital proxemics*. Environment was held constant across all sessions, making use of a large outdoor environment. However, the possibilities for environment design are one of the most exciting aspects of *digital proxemics*. Whyte's foundational work on the design of urban spaces [42] has already inspired similar approaches in virtual spaces [30]. Moore et al. [30] note the challenge of restraint and good design in the absence of physical constraints. Hillier and Hanson clearly note the problems of designing space with good intentions but poor assumptions: "For the first time, we have the problem of a 'designed' environment that does not 'work' socially, or even one that generates social problems that in other circumstances might not exist: problems of isolation, physical danger, community decay and ghettoisation." Without extending our understanding of environment design as a social-spatial place in the virtual, we risk creating spaces that cause more problems than they solve.

6 CONCLUSION

Designing effective experiences for virtual environments requires an intricate knowledge of how people make use of and behave in virtual spaces. We propose a research space for *digital proxemics* that addresses human activity, social signals, audio design, and environment to lay a foundation in this area. We completed a quantitative study exploring our initial hypotheses around audio designs with different levels of background noise and the impact of different social

[3]Blair MacIntyre designs audio differently based on expected group sizes. urlhttps://blairmacintyre.me/2020/04/03/vr2020-design-of-a-poster-room/

signals available to HMD and desktop PC users. Although this research only scratches the surface of the broader challenges, we hope this inspires future research into *digital proxemics* and improves how we collaborate and socialise in virtual environments in the future.

REFERENCES

[1] Mark S. Ackerman, Jack Muramatsu, and David W. McDonald. 2010. Social regulation in an online game: Uncovering the problematics of code. In *Proceedings of the 16th ACM International Conference on Supporting Group Work, GROUP'10*. 173–182. https://doi.org/10.1145/1880071.1880101

[2] Sun Joo (Grace) Ahn, Laura Levy, Allison Eden, Andrea Stevenson Won, Blair MacIntyre, and Kyle Johnsen. 2021. IEEEVR2020: Exploring the First Steps Toward Standalone Virtual Conferences. *Frontiers in Virtual Reality* 2, April (2021), 1–15. https://doi.org/10.3389/frvir.2021.648575

[3] Jeremy N. Bailenson, Jim Blascovich, Andrew C. Beall, and Jack M. Loomis. 2001. Equilibrium theory revisited: Mutual gaze and personal space in virtual environments. *Presence: Teleoperators and Virtual Environments* 10, 6 (2001), 583–598. https://doi.org/10.1162/105474601753272844

[4] Steve Benford, John Bowers, Lennart E. Fahlén, Chris Greenhalgh, and Dave Snowdon. 1995. User embodiment in collaborative virtual environments. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '95*. 242–249. https://doi.org/10.1145/223904.223935

[5] Steve Benford, Dave Snowdon, Andy Colebourne, Jon O'Brien, and Tom Rodden. 1997. Informing the design of collaborative virtual environments. *Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work* (1997), 71–79. https://doi.org/10.1145/266838.266866

[6] John Bowers, James Pycock, and Jon O'Brien. 1996. Talk and embodiment in collaborative virtual environments. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '96*. 58–65. https://doi.org/10.1145/238386.238404

[7] Philip Brey. 1999. The ethics of representation and action in virtual reality. *Ethics and Information Technology* 1, 1 (1999), 5–14. https://doi.org/10.1023/A:1010069907461

[8] Barry Brown and Marek Bell. 2004. CSCW at play : ' There ' as a collaborative virtual environment. In *CSCW 2004*. 350–359.

[9] E F Churchill and D Snowdon. 1998. Environments : An Introductory Review of Issues and Systems. *Virtual Reality* 3, 1 (1998), 3–15. https://link.springer.com/content/pdf/10.1007/BF01409793.pdf

[10] Nicolas Ducheneaut, Mh Wen, Nicholas Yee, Greg Wadley, Palo Alto, and Palo Alto. 2009. Body and Mind: A Study of Avatar Personalization in Three Virtual Worlds. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1151–1160. https://doi.org/10.1145/1518701.1518877

[11] Mike Fraser, Tony Glover, Ivan Vaghi, Steve Benford, Chris Greenhalgh, Jon Hindmarsh, and Christian Heath. 2000. Revealing the realities of collaborative virtual reality. In *Proceedings of the Third International Conference on Collaborative Virtual Environments*. 29–37. https://doi.org/10.1145/351006.351010

[12] Doron Friedman, Anthony Steed, and Mel Slater. 2007. Spatial social behavior in second life. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 4722 LNCS (2007), 252–263. https://doi.org/10.1007/978-3-540-74997-4{_}23

[13] Thomas Funkhouser, Nicolas Tsingos, and Jean-Marc Jot. 2003. Survey of Methods for Modeling Sound Propagation in Interactive Virtual Environment Systems. *Presence: Teleoperators and Virtual Environments, MIT Press* (2003), 1–53.

[14] Juan C. Gil-Carvajal, Jens Cubick, Sébastien Santurette, and Torsten Dau. 2016. Spatial Hearing with Incongruent Visual or Auditory Room Cues. *Scientific Reports* 6, November (2016), 1–10. https://doi.org/10.1038/srep37342

[15] Erving Goffman. 1959. *The Presentation of Self in Everyday Life*. 259 pages. https://doi.org/10.2307/2089106

[16] Erving Goffman. 1963. *Behavior in Public Places: Notes on the Social Organization of Gatherings*. 248 pages. https://doi.org/papers3://publication/uuid/4EEC120E-5776-413F-B43F-044C09251F5F

[17] Saul Greenberg, Nicolai Marquardt, Till Ballendat, Rob Diaz-Marino, and Miaosen Wang. 2011. Proxemic interactions. *Interactions* (2011). https://doi.org/10.1145/1897239.1897250

[18] A. Guye-Vuillème, T. K. Capin, I. S. Pandzic, N. Magnenat Thalmann, and D. Thalmann. 1999. Nonverbal communication interface for collaborative virtual environments. *Virtual Reality* 4, 1 (1999), 49–59. https://doi.org/10.1007/BF01434994

[19] Edward Hall. 1969. *The Hidden Dimension : man's use of space in public and in private*. 217 pages.

[20] Heiko Hecht, Robin Welsch, Jana Viehoff, and Matthew R. Longo. 2019. The shape of personal space. *Acta Psychologica* 193, April 2018 (2019), 113–122. https://doi.org/10.1016/j.actpsy.2018.12.009

[21] B. Hillier and J. Hanson. 1988. *The social logic of space*. https://doi.org/10.4324/9780429450174-9

[22] Tammo Houtgast and Herman JM Steeneken. 1971. Evaluation of speech transmission channels by using artificial signals. *Acta Acustica united with Acustica* 25, 6 (1971), 355–367.

[23] Adam Kendon. 1990. *Conducting interaction. Patterns of behaviour in focused encounters*. 308 pages. http://books.google.com/books?id=7-8zAAAAIAAJ

[24] Duc Anh Le, Blair MacIntyre, and Jessica Outlaw. 2020. Enhancing the Experience of Virtual Conferences in Social Virtual Environments. In *Proceedings - 2020 IEEE Conference on Virtual Reality and 3D User Interfaces, VRW 2020*. 485–494. https://doi.org/10.1109/VRW50115.2020.00101

[25] Jie Li, Vinoba Vinayagamoorthy, Julie Williamson, David A. Shamma, and Pablo Cesar. 2021. *Social VR: A New Medium for Remote Communication and Collaboration*. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3411763.3441346

[26] Joan Llobera, Bernhard Spanlang, Giulio Ruffini, and Mel Slater. 2010. Proxemics with multiple dynamic characters in an immersive virtual environment. *ACM Transactions on Applied Perception* 8, 1 (2010). https://doi.org/10.1145/1857893.1857896

[27] Antonella Maselli and Mel Slater. 2013. The building blocks of the full body ownership illusion. *Frontiers in Human Neuroscience* 7, March (2013), 1–15. https://doi.org/10.3389/fnhum.2013.00083

[28] John A. McArthur. 2016. *Digital Proxemics.* https://doi.org/10.3726/978-1-4539-1724-4

[29] Joshua McVeigh-Schultz, Anya Kolesnichenko, and Katherine Isbister. 2019. Shaping Pro-Social Interaction in VR. In *CHI '19*. 1–12. https://doi.org/10.1145/3290605.3300794

[30] Robert Moore, E. Hankinson Gathman, and Nicolas Ducheneaut. 2009. From 3D space to third place: The social life of small virtual spaces. *Human Organization* 68, 2 (2009), 230–240. https://doi.org/10.17730/humo.68.2.q673k16185u68v15

[31] Alexander Muryy and Andrew Glennerster. 2021. Route selection in non-Euclidean virtual environments. *PLoS ONE* 16, 4 April (2021), 1–23. https://doi.org/10.1371/journal.pone.0247818

[32] B Nardi and J Harris. 2010. Strangers and friends: Collaborative play in World of Warcraft. In *International Handbook of Internet Research*. 395–410. https://doi.org/10.1007/978-1-4020-9789-8

[33] Carman Neustaedter and Elena Fedorovskaya. 2009. Presenting identity in a virtual world through avatar appearances. In *Proceedings of Graphics Interface 2009*. 183–190. http://portal.acm.org/citation.cfm?id=1555921

[34] Iana Podkosova and Hannes Kaufmann. 2018. Co-presence and proxemics in shared walkable virtual environments with mixed colocation. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology, VRST*. https://doi.org/10.1145/3281505.3281523

[35] Daniel Pohl and Markus Achtelik. 2019. Personalized personal spaces for virtual reality. *26th IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2019 - Proceedings* (2019), 1128–1129. https://doi.org/10.1109/VR.2019.8797773

[36] Bill Rogers, Masood Masoodian, and Mark Apperley. 2018. A virtual cocktail party: Supporting informal social interactions in a virtual conference. *Proceedings of the Workshop on Advanced Visual Interfaces AVI* (2018). https://doi.org/10.1145/3206505.3206569

[37] Ralph Schroeder, Ilona Heldal, and Jolanda Tromp. 2006. The usability of collaborative virtual environments and methods for the analysis of interaction. *Presence: Teleoperators and Virtual Environments* 15, 6 (2006), 655–667. https://doi.org/10.1162/pres.15.6.655

[38] Theresa Jean Tanenbaum, Nazely Hartoonian, and Jeffrey Bryan. 2020. "how do i make this thing smile?": An Inventory of Expressive Nonverbal Communication in Commercial Social Virtual Reality Platforms. *Conference on Human Factors in Computing Systems - Proceedings* (2020), 1–13. https://doi.org/10.1145/3313831.3376606

[39] David Thery and Brian F. G. Katz. 2021. Auditory perception stability evaluation comparing binaural and loudspeaker Ambisonic presentations of dynamic virtual concert auralizations. *The Journal of the Acoustical Society of America* 149, 1 (2021), 246–258. https://doi.org/10.1121/10.0002942

[40] Matteo Varvello, Stefano Ferrari, Ernst Biersack, and Christophe Diot. 2011. Exploring second life. *IEEE/ACM Transactions on Networking* 19, 1 (2011), 80–91. https://doi.org/10.1109/TNET.2010.2060351

[41] Alessandro Vinciarelli, Hugues Salamin, Anna Polychroniou, Gelareh Mohammadi, and Antonio Origlia. 2012. From nonverbal cues to perception: Personality and social attractiveness. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 7403 LNCS (2012), 60–72. https://doi.org/10.1007/978-3-642-34584-5{_}5

[42] William H. Whyte. 1982. *The Social Life of Small Urban Spaces*. Vol. 10. 466–468 pages. https://doi.org/10.1177/089124168201000411

[43] Laurie M. Wilcox, Samuel Elfassy, Cynthia Grelik, and Robert S. Allison. 2006. Personal Space in Virtual Reality. *ACM Transactions on Applied Perception* 3, 4 (2006), 412–428. https://doi.org/10.1145/1190036.1190041

[44] Julie Williamson, Jie Li, Vinoba Vinayagamoorthy, David A. Shamma, and Pablo Cesar. 2021. Proxemics and Social Interactions in an Instrumented Virtual Reality Workshop. In *ACM Conference on Human Factors in Computing Systems (CHI)*. 1–20. https://doi.org/10.1145/3411764.3445729

[45] Noelle L. Wood and Nelson Cowan. 1995. The Cocktail Party Phenomenon Revisited: Attention and Memory in the Classic Selective Listening Procedure of Cherry (1953). *Journal of Experimental Psychology: General* 124, 3 (1995), 243–262. https://doi.org/10.1037/0096-3445.124.3.243