Multiplex Vision: Understanding Information Transfer and F-Formation With Extended 2-Way FOV

Mark Armstrong
Keio University Graduate School of Media Design
Tokyo, Japan
mark@keio.jp

Keitaro Tsuchiya
Keio University Graduate School of Media Design
Tokyo, Japan
k.tsuchiya@kmd.keio.ac.jp

Feng Liang
Ignition Point, Inc
Tokyo, Japan
feng.liang@ignitionpoint-inc.com

Kai Kunze
Keio University Graduate School of Media Design
Tokyo, Japan
kai@kmd.keio.ac.jp

Yun Suen Pai
The University of Auckland
Auckland, New Zealand
yspai1412@gmail.com

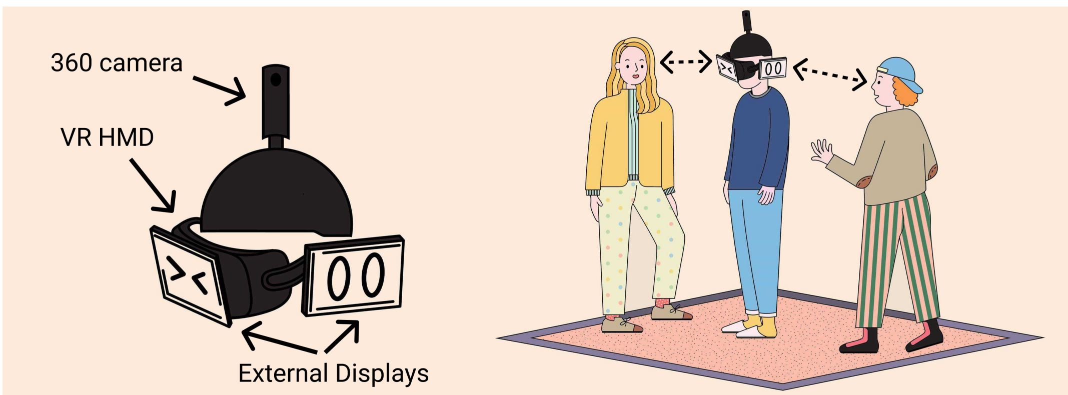

Figure 1: The Multiplex Vision prototype built using a 360° camera, with VR HMD and 3 LCD displays mounted externally (left). We illustrate how it can be used to facilitate multidirectional conversation, where (right) the user is able to maintain eye contact with multiple participants simultaneously.

ABSTRACT

Research in sociology shows that effective conversation relates to people's spatial and orientational relationship, namely the proxemics (distance, eye contact, synchrony) and the F-formation (orientation and arrangement). In this work, we introduce novel conversational paradigms that effects conventional F-formation by introducing the concept of multi-directional conversation. Multiplex Vision is a head-mounted device capable of providing a 360° field-of-view (FOV) and facilitating multi-user interaction multi-directionally, thereby providing novel methods on how people can interact with each other. We propose 3 possible new forms of interactions from our prototype: one-to-one, one-to-many, and many-to-many. To facilitate them, we manipulate 2 key variables, which are the viewing parameter and the display parameter. To gather feedback for our system, we conducted a study to understand information transfer between various modes, as well as a user study on how different proposed paradigms effect conversation. Finally, we discuss present and future use cases that can benefit from our system.

VRST '20, November 1–4, 2020, Virtual Event, Canada

ACM ISBN 978-1-4503-7619-8/20/11…$15.00
https://doi.org/10.1145/3385956.3418954

CCS CONCEPTS

• **Applied computing → Sociology**; • **Human-centered computing → Mixed / augmented reality**.

KEYWORDS

360 field-of-view, vision augmentation, F-formation, conversation

ACM Reference Format:
Mark Armstrong, Keitaro Tsuchiya, Feng Liang, Kai Kunze, and Yun Suen Pai. 2020. Multiplex Vision: Understanding Information Transfer and F-Formation With Extended 2-Way FOV. In *26th ACM Symposium on Virtual Reality Software and Technology (VRST '20), November 1–4, 2020, Virtual Event, Canada.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3385956.3418954

1 INTRODUCTION

Face to face communication is a crucial way of transferring knowledge and sharing our feelings. When we are talking to each other, we also rely heavily on nonverbal clues, such as body language, eye contact, etc. For effective face-to-face communication to occur, sociological theory suggests that the distance between people, or proxemics [11], and the way they orient each other, or F-formation [15], greatly influences this. This will then eventually affect other key considerations like body language, eye contact, and so on.

We wish to present a method that may contribute towards human interaction itself by having a device that is capable to create new potential interaction modalities by introducing novel F-formations. With this, we present a prospective notion; with the aid of HCI technology, can we complement, modify, or completely introduce novel methods for humans to interact with each other if we were to eliminate a key requirement for an effective interaction, namely the directional cue and/or orientation, while keeping the required distance (therefore, not considering devices for long distant communication)?

In this work, we propose Multiplex Vision as shown in Figure 1, which is a device that presents new paradigms for people to interact with each other. It couples the use of a 360° camera to allow the change in environmental viewing for the user, with displays around the user's head for additional information sharing. Based on these two parameters, we are able to change both the viewing and display modality allowing us to introduce novel multi-directional paradigms for human interaction. We propose three interaction paradigms: one-to-one, one-to-many and many-to-many. This work is overall motivated by simply exploring the possibility of multi-directional interaction, yet we see some good potential applications that may arise from this, such as a tool for disabled people, or for surveillance. Our research goals are threefold:

(1) To develop a prototype that can facilitate multi-directional conversation.
(2) To introduce new conversational paradigms with the proposed prototype.
(3) To perform exploratory studies on the effects of Multiplex Vision on conversation across different proposed paradigms.

2 RELATED WORKS

In this section, we present work related to Multiplex Vision by first understanding the definition of proxemics and F-formation, followed by how vision augmentation and wearable displays play a role in developing Multiplex Vision.

2.1 Proxemics and F-Formation

Proxemics, according to Hall [11], is defined as the study of spaces between humans. In sociology, the distance between two individuals is categorized as intimate, personal, social, and public space. The regular distance between two individuals conversing falls into social space, between 1.2m to 2.1m. Proxemics have been a key consideration in HCI, as it allows researchers to define digital play experiences [21], interaction with artifacts [8], public guidance systems [30], human-human and human-robot interaction [22, 27, 31]. Instead of just distance, Dostal et al. [3] also expanded on proxemics by investigating the horizontal and vertical angles from a display and how the display content changes based on these parameters. Proxemics has been associated with other implicit interactions like nonverbal cues and body orientation, but it is generally defined by interpersonal physical distance [18].

F-formation, also known as facing formation on the other hand, according to Kendon [15], arises when two or more people sustain a spatial and orientational relationship where the space between them is accessible in an equal, direct and exclusive way. Other fields of research have benefited from the understanding of this, such as in human-robot interaction where a robot is designed to fluidly participate in human conversation [12]. The orientation of the participants directly affects the eye contact as well, which is another important factor in communication. In the research conducted by Rychlowska et al. [25], results from their studies supported that embodied simulation can be triggered by eye contact. F-formation was overall meant for a stationary conversation scenario. In computer science, F-formations are often used either for detection [5, 13] or for analysing the spatial patterns [19]. For designing interaction modalities, Marquardt et al. [18] analyzed micro-mobility, which is the fine-grained orientation of physical artifacts. In our work, however, we essentially integrate the concept of micro-mobility with F-formation, where the artifact is mounted on the user's head.

2.2 Vision Augmentation

There are a lot of approaches to improve the efficiency of visual navigation in Virtual, Augmented and Mixed Realities [9, 24]. for brevity we just focus on vision augmentation. Human vision augmentation is often designed to circumvent human limitations by increasing the FOV, spatial perception, and so on. For example, SpiderVision developed by Fan et al. [4] proposes a wearable device to extends the human FOV by augmenting one's back view. Though this solution does not allow full 360° view to the user at all times, it inspired the implementation for our Layered View mode. Unconstrained Neck [26] was another work that expanded on human FOV using a non-linear mapping of the neck movement with the view rotation, allowing users to see behind them easily. There have also been other hardware-based approaches, such as by Grunefeld et al. [9] and Orlosky et al. [23] that used an LED ring around the lenses and a stereoscopic fisheye lens respectively to increase FOV. The only related work we found that provides this is FlyViz [1] and PanoFlex [16]. The former used a 360° camera with an HMD to show an equirectangular view of the environment to the user, whereas the latter uses the same hardware, but instead proposed a dynamic distortion solution to prioritize a user's direction-of-interest.

Our work instead proposes vision augmentation not specifically to increase FOV, but as a method to understand user behavior when provided a device that can facilitate multi-directional conversation. We analyze user feedback and behavior similar to Parallel Eyes [14] that explored human behavior when they are suddenly introduced with the capability to view from four points-of-views simultaneously.

2.3 Wearable External Displays

A conversation is a two-way process, and how others see us is as important as how we see them. Looking at some previous works on how displays are attached externally, FrontFace presented by Chan

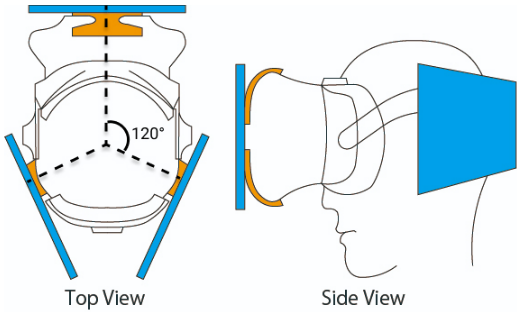

Figure 2: Overview of the hardware. The light blue area shows the display and the orange area shows the 3D-printed mounting parts to connect the display to the HTC Vive Pro Eye HMD.

et al. [2] used the external display, which is essentially a smartphone, to show the user's gaze direction when using a VR device to outsiders. This is followed up by TransparentHMD proposed by Mai et al. [17] which used a smartphone display to not only show the gaze point, but the entire face of the user which is normally occluded by the HMD. Instead of reflecting the user's face, ChameleonMask proposed by Misawa et al. [20] is instead used as a telepresence tool where a surrogate user mounts the face of a remote user on theirs via a HMD. On the interaction side, FaceDisplay proposed by Gugenheimer et al. [10] uses 3 touch displays and a depth camera attached on the HMD to facilitate interaction between the VR and non-VR user. Most of these works are specifically catered towards the VR audience and are meant to bridge the gap between VR and non-VR. In this regard, our work is aimed more towards the gap between users and non-users of vision augmentation tools, where the display is wrapped around the user to facilitate multi-directional communication.

3 PROTOTYPE IMPLEMENTATION

This section will discuss the proposed hardware and software to build Multiplex Vision.

3.1 Hardware

Multiplex Vision utilizes a commercial VR device, in this case the HTC Vive Pro Eye, to display the content to the user. This VR device was chosen because of its ability to record eye tracking data, which allowed the implementation of a tracking reticle for recording purposes, as well as for developing one of our display modes. The rig consists of a helmet fitted with a mount on top for a 360° camera. In this prototype, the RICOH Theta V FullHD 360° camera was used for live tethered streaming of the 360° content. A display (resolution of 1200 x 800) is attached to the front of the headset, providing a forward facing video stream for observers to watch. Accompanying this are two additional displays mounted around the HMD itself, at an angle of 120° and -120° respectively, illustrated in Figure 2. These displays serve as a multi-directional video stream that can be seen from wherever the observer chooses to stand.

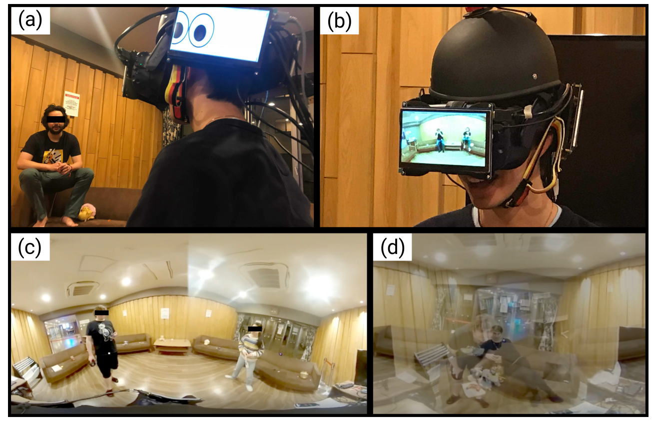

Figure 3: The developed modes for vision and display manipulation: (a) Gaze Share and (b) View Share shows the developed vision manipulations, whereas (c) Equirectangular View and (d) Layered View shows the developed display manipulations

Video from the 360° camera is streamed to the user as well as the internal engine which handles the formatting of data to the attached LCD Displays. The data stream was sent through a bundle of 4 micro-USB cables as well as 4 HDMI/DVI cables held together via zip ties. The display holder and mount for attaching the display to the HMD were made by 3D printing. The front display was created by modifying a 3D model of an existing mount for other parts. The side displays are attached by a 3D printed mount that is inserted into the side of the head of the HMD's band.

3.2 Vision Manipulation

We developed two variations; equirectangular view and layered view shown in Figure 3 (c) and (d) respectively. For the equirectangular view, we map the direct video feed from the 360° camera mounted on the head to a plane instead of the traditional sphere around the user in the virtual environment. This provides the user with a complete 360° rectangular HUD at all times. To our knowledge, the only other related works that provide complete 360° vision is FlyVIZ [1] and PanoFlex [16]. This novel feature allows the user to monitor the surroundings at all time and also be aware of the presence of other people and their facing directions. Using the built-in eye trackers, we then track the gaze point of the user when looking at the equirectangular view. This information will later be used for the display manipulation.

The second developed viewing mode is the layered view. This view consists of 2 viewing planes layered back to back in front of the user. After initial testing, we found that an opacity of 70% for the planes allows the user to see both views at the same time comfortably. Furthermore, overlap in the content for both views need to be avoided to minimize confusion. Each of the viewing planes are therefore a 170° FOV to the left and right of the user with 0 overlap. This effectively leaves a blind-spot of 10° at the front and back of the user. However from our testing, we found that despite the existence of the blind-spots, 10° is a relatively small angle to

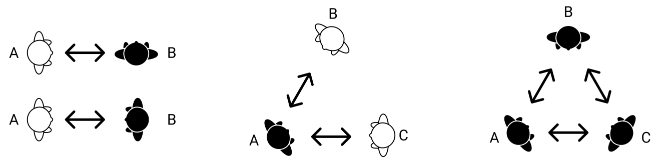

Figure 4: Proposed paradigms for (left) multidirectional one-to-one, (middle) one-to-many with the ability to maintain eye contact two other participants, and (right) many-to-many where all participants are users of Multiplex Vision.

the point that the user just needs to slightly turn their head to be able to see the front and back direction. An example related work that uses layered view is SpiderVision [4], which overlays the front and back view. We chose the left and right view instead so that we can map this to the corresponding external displays.

3.3 Display Manipulation

We developed two variations; view sharing and gaze sharing shown in Figure 3 (b) and (a) respectively. For the view sharing mode, we have the external displays showing the direction of what the user is looking at, mapped with spherical projection. This mode attempts to recreate the feeling of a remote video call session (the ability to see themselves in a conversation), but is instead directly mounted on a user who has 360° vision. Using this method, the non-user will be able to see what the user is looking at, for ease of information transfer. This mode was developed to aid conversational flows when needing to describe about the presence of physical objects at a specific direction, which can be viewed on the display when the user simply looks at it. Furthermore, when maintaining eye contact, view sharing will show the non-user their own face since the user is looking at them, allowing for self evaluation, akin to seeing our own faces when conversing with a video call.

For the gaze sharing mode, we instead project the gaze point of the user onto the display. This mode attempts to replicate how a conventional face-to-face conversation would be when a participant has a 360° vision. The external displays serve as extended displays that goes all around the user. Eye blinks are also portrayed accurately when the system detects them. Depending on the user's viewing mode, the gaze point, illustrated as two virtual eyes, are shown on the display. For example, if the user is viewing the environment in equirectangular mode, the external display will accurately show the gaze point, i.e if the user is looking left, the eyes will appear on the left. This method allows the user to convey gaze information multi-directionally.

4 F-FORMATION PARADIGMS

In this section, we propose new F-formation paradigms with the use of Multiplex Vision.

4.1 Paradigm 1: One-to-One

In a one-to-one communication, it involves two individuals each acting as both the speaker and the listener. In the first proposed scenario we have person A talking to person B who has Multiplex Vision (Visualized in Figure 4(left) as the black avatar). With this, Person B can simply look in any direction while still carrying on the conversation. Person A would still be able to maintain eye contact, independent of person B's facing direction by having the display facing Person A showing Person B's gaze point. For this to work, Multiplex Vision utilizes the equirectangular view coupled with gaze sharing. If the scenario of the conversation requires person B to describe an object of reference, or to show a physical object within the vicinity, person B is also able to transfer the information using the external display's view sharing mode.

4.2 Paradigm 2: One-to-Many

For a one-to-many scenario, we introduce a group conversation where only one user of Multiplex Vision is present. We illustrate this in Figure 4(middle) with the black avatar as the user.

When combining equirectangular view with view sharing, the conversation becomes multi-directional, as in the one-to-one paradigm. However, the view mode that shares the user's viewpoint enables other participants to see and acknowledge the flow of conversation. If participant B sees that participant A is looking towards himself/herself, the flow of conversation is clearly directed to participant B, which is made aware to participant C too. Likewise, the same can be said if participant A looks at participant C.

When combining layered view with gaze sharing, the device is then able to project the gaze point of the user onto both participants B and C. With the use of the layered view, participant A can effectively look at both B and C at the same time i.e maintaining eye contact with both participants, which cannot be achieved without the use of Multiplex Vision. This combination of modes is more suitable if participant A plays a more major role as a speaker compared to listener, with scenarios like giving a speech or presentation.

4.3 Paradigm 3: Many-to-Many

For a many-to-many communication scenario, we look at all participants being the user of Multiplex Vision. Each participant would be able to choose between a variety of developed modes for viewing and displaying to others, illustrated in Figure 4(right). That way, we achieve true flexibility in multi-directional conversation, where each participant may freely face any direction. We propose such an interaction to be possible in any form of grouped gathering without a prioritised speaker or listener.

5 STUDY 1: INFORMATION TRANSFER

The goal of our first study is to evaluate the capability and effectiveness of information transfer using our system based on a simple search task. A participant will try to look for objects placed in the nearby vicinity based on the information gathered from the external display. The mode combinations that will be tested are the following: equirectangular + view sharing, equirectangular + gaze sharing, and equirectangular + no displays. The independent variable for this study is therefore the display modes (nothing displayed, gaze share and view share), whereas the dependent variable is the time required to complete the search task. We also employ a "think aloud" protocol, whereby participants at anytime are free to voice their thoughts and provide feedback. We hypothesize that the use of Multiplex Vision, particularly View Sharing, should provide

Figure 5: Setup of Study 1 with the Lego pieces places around the user who will be sitting at the middle

the most information to the listener in a search task, as the listener sees exactly what the speaker sees.

5.1 Participant

We recruited 8 participants between the ages of 19 to 23 (2 Females, Mean: 20.75, SD: 1.49) to participate in this within-subject study. Additionally, all participants were students and have never experienced VR before except for one who only spent a brief amount of time watching videos in VR. For this study however, they will not be using the device, but instead is required to communicate with the user.

5.2 Procedure

Prior to the experiment, we have each of the participants fill a consent form stating that there will be no compensation and they are free to end the experiment at any time if they wish to do so. Due to the conversational nature of the experiment, we require a participant and an experiment conductor to join the study at the same time, where the role of the participant is the listener (non-user), and the conductor as the speaker (user). We use a conductor as the speaker to ensure the information is delivered the same for each condition. After we brief them regarding the procedure, the speaker sits in the middle of the room, where 26 Lego pieces are scattered around him/her in multiple directions but are within line of sight. There are also 13 random objects placed around the room in predetermined positions to either distract or assist the participant in their search task. The participant is free to stand anywhere in the room. When the study begins, the conductor is prompted in the VR display a description of a distinct Lego piece. The conductor will then convey the information to the listener based on a script, and will look in the direction of the Lego (using the random objects as reference points). As soon as the Lego description is finished being spoken, a timer is started, and we time how long it took for the participant to complete the task by placing the Lego in the conductor's hand. If the participant is confused and asks if he/she have found the correct piece, the conductor may confirm or deny the correctness of the acquired Lego. We continue to perform this task for 10, non-repeated Legos. The sequence of the Legos are

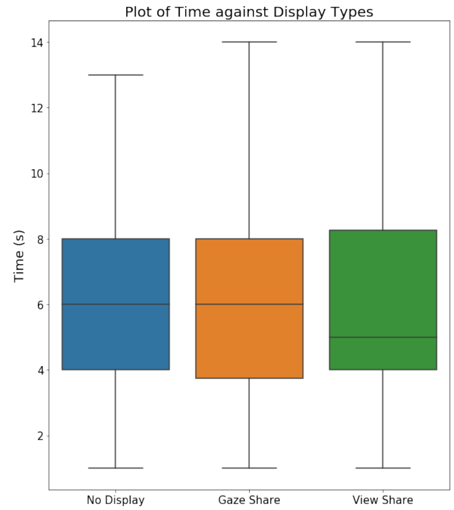

Figure 6: Results for Study 1

randomized for each of the modes. At the end of the session, we reset the Legos' positions and proceed to the next mode in Latin Square order with a 5 minute break between each mode.

5.3 Results and Feedback

The gathered results for Study 1 are shown on Figure 6. It can be seen that the time spent for the participant to seek out the correct Lego pieces for each of the conditions are comparatively similar. From a Friedman test, there was no statistically significant difference in the time spent depending on the modes (x^2 = 0.063, p = 0.96). The average time taken is 7.2s, 7.1s and 7.5s for no display, gaze share and view share respectively.

P1 and P2 mentioned that even though the display was more of a good initial pointer for the location, it eventually felt distracting and they would rather just focus the search on the environment instead. P3 was convinced that the gaze share mode was the most helpful, even though there was a slight delay. The delay was not due to the system itself, but rather from the OBS screen recording software running at the same time (from testing, the system typically runs at 40 frames per second without screen recording). However, P7 mentioned that gaze share was difficult to rely on for exact locations, whereas view share was initially confusing. P6 was observed to mostly use the gaze share mode for verification after locating the Lego piece. Similar observation was obtained from P8 who actually relied quite strongly on the view share mode, to the point that he/she misunderstood the color of the Lego and decided to first prioritise what is shown on the display before selection. It was found that most participants ended up simply memorizing the locations of the Lego pieces over time. Additionally, it was observed that for gaze share, the participants seemed to get an understanding of which direction to look mostly just by watching which direction the eyes started to move. Once they looked in that general direction, they could rely on their own intuition to find the object.

Table 1: Proposed 2x2 study condition

User Type / Paradigm	User (Equirectangular / Layered View	Non-User (Gaze/View Share)
1-1	Condition 1	Condition 2
1-many	Condition 3	Condition 4

Table 2: Modality Questionnaire

Q#	Statement
Q1	The device was an acceptable medium for conversation
Q2	The device felt like a comfortable way to converse
Q3	The device provided a convenient way for conversing without caring about facing direction
Q4	The device was as effective as if the speaker was actually facing me
Q5	The device is a medium I would use again for conversation

6 STUDY 2: CONVERSATIONAL SCENARIOS

The goal of this study is to evaluate how the use of Multiplex Vision can actually effect a real face-to-face communication, either between two or more individuals. The participants test the proposed paradigms and provide their feedback regarding the modality [28] and perceived quality of communication [6, 7] shown on Table 2 and 3 using a slightly modified version of the questionnaires from the cited work to be used in our context. The independent variable for the study is the role of the participant and the formation type, whereas the dependent variables are the aforementioned questionnaire results, as shown in Table 1. Like Study 1, we also employ a "think aloud" protocol, whereby participants at anytime are free to voice their thoughts and provide feedback. We hypothesize that there would be no significant differences in perceived communication quality between the conditions, but rather an implicit change in communication behavior.

Study 2 was conducted back-to-back with Study 1, and so we recruited the same participants who completed the Study 1. Study 2 was performed after a 10 minute rest from the completion of Study 1. Like Study 1, the participant will be interacting with the experiment conductor for each of the proposed paradigms. In two cases, one conductor, and in two cases, two conductors. The paradigm sequence is also arranged in a Latin Square order to minimize ordering effect, with a 5 minute break between each paradigm.

6.1 Procedure

After we brief the participant about the procedure, the speaker sits in the middle of the room, with the listener facing the speaker. Shown on Table 1, each paradigm will have two modes for the participant to try out depending on their role as either the user or non-user. A 5 minute conversation is initiated (2.5 minutes per mode). Four generic and easily conversible topics of conversations are assigned, which are about effects of the coronavirus, desired superpowers, countries to travel, and preferred food. The facing directions are also predefined for each paradigm at the start of the experiment. However, after the conversation starts, participants are free to move around as we observe and record their behavioural response, as long as they stay within the VR tracking zone. The selected paradigms are 1) one-to-one as user, 2) one-to-one as non-user, 3) one-to-many as user, and 4) one-to-many as non-user. Many-to-many was not tested because that would require multiple prototypes, of which we only have one. Since all of the proposed paradigms are based on conversations, we have at least one of the experiment conductor participating as well.

Table 3: Perceived Quality Questionnaire

Q#	Statement
Q1	I could readily tell when my partner(s) was listening to me
Q2	I was able to take control of the conversation when I wanted to
Q3	It was easy for me to contribute to the conversation
Q4	The conversation seemed highly interactive
Q5	There were not many frequent and inappropriate interruptions
Q6	This felt like a natural conversation
Q7	I found it easy to keep track of the conversation
Q8	I felt completely absorbed in the conversation
Q9	I had a real sense of personal contact with my conversation partner(s)
Q10	I was very aware of my conversation partner(s)
Q11	My partner(s) was friendly
Q12	My partner(s) took a personal interest in me
Q13	I trusted my partner(s)
Q14	I enjoyed talking to my partner(s)
Q15	I would be interested in meeting my partner(s) face-to-face

For the 1) one-to-one as user paradigm, the participant uses Multiplex Vision, whereas the conductor plays the role of the non-user in the conversation. At the start, the user does not face the non-user, whereas the non-user faces one of the user's side displays during conversation. Since the participant will be evaluated as the user, we provide 2 viewing modes for this paradigm: equirectangular and layered view. The external displays are set to gaze mode by default so that the non-users can maintain eye contact with the participant.

For the 2) one-to-one as non-user paradigm, the role is simply reversed, allowing the evaluation of a non-user when facing the user on one of the external displays during conversation. Since the participant will be evaluated as the non-user, the view sharing and gaze sharing modes are provided. The user view is set to equirectangular, which we treat as the default viewing mode, as it is needed to reflect the user's gaze and/or view points for the participant to see on the display.

For the 3) one-to-many as user paradigm, the participant acts as the user of Multiplex Vision while talking to two non-users in a group conversation. At the start, the non-users each face the side displays of the user. The provided viewing modes are therefore the same as condition 1, with the display mode set to gaze share by

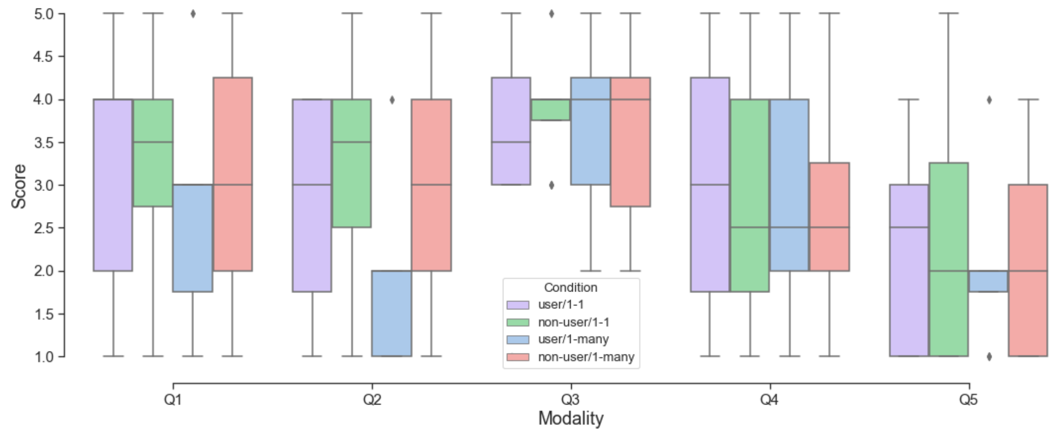

Figure 7: Results for Modality Questionnaire from Table 2

Figure 8: Example of Study 2, where the non-user is free to comfortably chat with the user as we observe the behaviour changes across four conditions.

default. This paradigm allows the evaluation of a group conversation experience as well as feedback towards the use of layered view when required to look at multiple people at the same time.

For the 4) one-to-many as non-user paradigm, one of the participant acts as the non-user, whereas the conductors act as the user and another non-user respectively. The provided display modes are therefore the same as condition 2, with the view mode set to equirectangular by default.

6.2 Results and Feedback

The gathered results for the modality questionnaire in Table 2 is shown in Figure 7. We can observe that for each of the questions across all four conditions, Q3 (the device is convenient) garnered the highest average score of 3.5, 3.9, 3.8 and 3.6 for each of the conditions respectively. The only average score below 2.5 is Q2 for condition 3 (average score of 1.9) and all conditions for Q5 (2.3, 2.4, 2 and 2.1 per condition). Q2 looked into determining if the device was comfortable to use while conversing, where we understand that the score of the participant playing the role of the user to be low mainly due to the weight of the device. Additionally, condition 3 required the participant to also converse in a group, which could potentially feel more challenging. Looking at Q5, participants think that the chance of them using the device for future conversation across all four conditions to be low. This could be attributed to several factors, including the comfort of using it as well as the short usage duration where they may not be accustomed to this method of conversing yet. However, the results for the perceived quality of communication are generally favourable, as shown in Figure 9. We first performed a Shapiro-Wilk test and found the results to not be normally distributed. To investigate if there are any interaction effects, we then perform an aligned rank transform (ART) [29] followed by a two-way ANOVA analysis. However, no significant interaction or main effect was present, showing us that conversations were still effective and felt relatively natural.

Analyzing the feedback, most of the participants initially think, the device was slightly intimidating, yet it was actually "convenient to converse without caring about the facing direction" (quoted from P3). This was due to both the enhanced FOV as well as how it is conveyed to the audience. Many participants also favour the equirectangular view over the layered view, stating that the layered view was distracting and difficult to focus. It was observed that, even though participants were given freedom to face any direction or to move around, they generally stayed stationary in a comfortable position (see Figure 8) and faced any one of the displays when conversing.

P2 mentioned that it was initially strange to converse with a person wearing a head-mounted display during condition 1, and required some time to get used to. P2 also much preferred equirectangular over layered view during condition 3, stating that the layered view was uncomfortable. For conditions 2 and 4, P4 found that gaze share was interesting to look at as the non-user since it looks and blinks like a real pair of eyes. For condition 1, P4 also mentioned that the layered view felt "trippy", whereas equirectangular view was surprisingly comfortable to use and more natural than initially perceived. This was reinforced in condition 3, where P4 also mentioned that layered view made it difficult to indicate the directions, as "turning the head does not feel like turning the head at all." P5

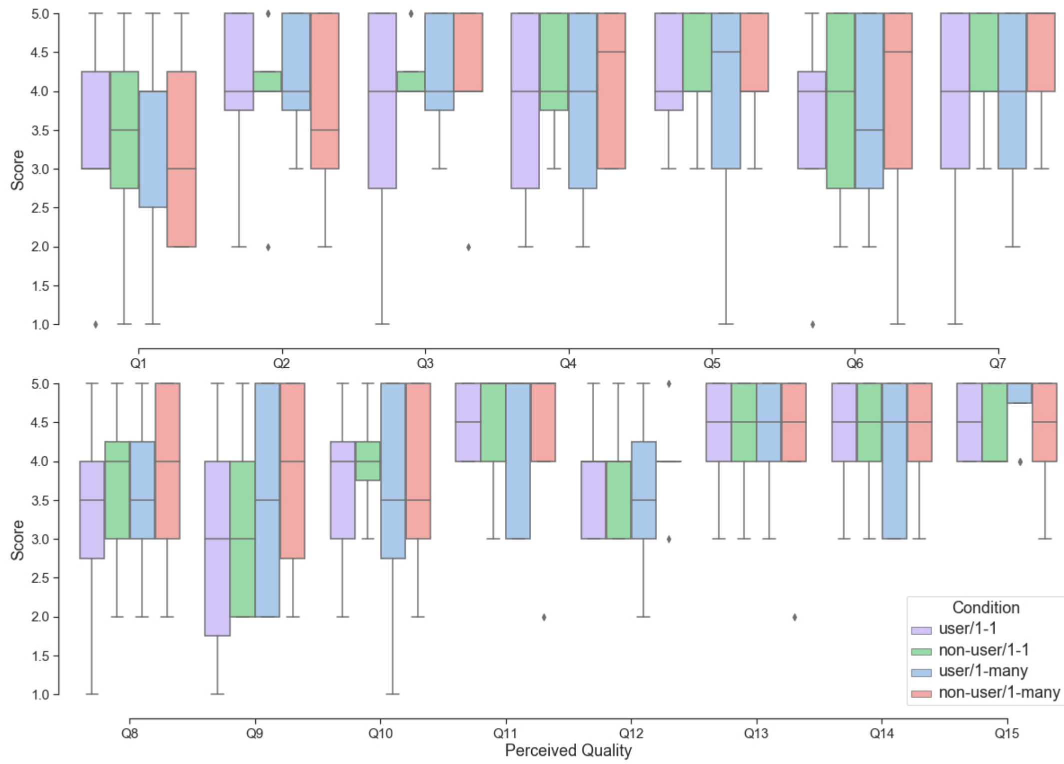

Figure 9: Results for Perceived Quality Questionnaire from Table 3

mentioned that equirectangular view was more convenient and easier to communicate with when compared to layered view. P5 also felt it distracting to use layered view in condition 3, especially if the others moved around. As a non-user for condition 4, P5 mentioned that "I only see the eyes moving without the head, so it feels less interactive." P8 mentioned that it blurs the line between the conventional use of video call with face-to-face conversation, by stating "it felt a bit like talking a very realistic non-player character (NPC) in a game" which we find to be interesting. This was for both condition 1 and 2, meaning as both the user and non-user for a one-to-one conversation. For condition 3 which was the last condition for P5, he/she stated that "I was getting used to the device already though I still feel dazed from layered view. The equirectangular view though, was like switching from a dream to reality." From the gathered feedback, layered view was overall less preferable for several key reasons; it felt more uncomfortable possibly due to difficulty in focusing on one thing at a time, directional cues become harder to perceive, and that any moving object present can be hard to differentiate between the layers.

In general, participants remarked that it felt a bit like "talking to a machine", which diminished the genuine feeling of a face-to-face interaction. The use of virtual eyes therefore, could possibly be further improved, either by using more realistic-looking representation, or a future device that captures the image of the user's eyes and project them on the screens. In this study, there was no mention of any uncanny valley effects, though rendering realistic eyes may cause this.

7 LIMITATIONS AND FUTURE WORK

It is still unclear how Multiplex Vision affects a person's sociological behaviour when they do not actually need to face someone when conversing with them. The social acceptability of using Multiplex Vision should also be discussed. In social gatherings, a certain level of contact and interactivity is expected from each participant. From our study results, it seems that the convenience of multi-directional viewing instead may limit physical movement like turning around.

Despite the existence of the 360° display, the voice projected from the user is not surround sound and is dependent on the user's facing direction, which will also directly influence the effectiveness of a face-to-face communication. Furthermore, due to the reliance on cameras and displays to achieve this work, the image resolution will undoubtedly affect the conversation.

The hardware prototype needs to also be much smaller and lighter for this to ever be considered as a ubiquitous device for now. As we are using commercially available VR HMDs for Multiplex Vision, future iterations of the device would benefit the form factor of our implementation. With multiple devices present, we can then proceed to evaluate Paradigm 3, alongside other novel viewing and display modes. We also plan to evaluate the possibilities of motion sickness and perceived workload for the new form factor, as it is currently not within the scope of this work.

8 ADDITIONAL APPLICATION CASES

Whereas related works target the VR audience by facilitating assymetrical interaction with those outside of VR [2, 10], we envisioned the use of Multiplex Vision towards a more generic, conversational use case. In this section, we look into other potential applications that can potentially benefit from Multiplex Vision.Since Multiplex Vision allows user to see the surroundings and interact with people around without turning their head, people with disability like spinal injury may benefit from this device.It can also be used as a security tool for both monitoring or increased spatial awareness. Lastly, without the need to actually face each other in a conversation, the phrase "say it, don't spray it" is rendered irrelevant.

9 CONCLUSION

We present Multiplex Vision, introducing new ways on how people interact with each other face-to-face using 360° FOV and surrounding external displays. We discuss on how it affects our understanding of F-formation, by proposing novel paradigms for augmented individuals to communicate face-to-face without considering facing orientation.

ACKNOWLEDGMENTS

This research is a collaboration between Keio Graduate School of Media Design, University of Auckland, and Ignition Point, Inc.

REFERENCES

[1] Jérôme Ardouin, Anatole Lécuyer, Maud Marchal, Clément Riant, and Eric Marchand. 2012. FlyVIZ: a novel display device to provide humans with 360 vision by coupling catadioptric camera with hmd. In *Proceedings of the 18th ACM symposium on Virtual reality software and technology*. 41–44.

[2] Liwei Chan and Kouta Minamizawa. 2017. FrontFace: facilitating communication between HMD users and outsiders using front-facing-screen HMDs. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–5.

[3] Jakub Dostal, Per Ola Kristensson, and Aaron Quigley. 2013. Multi-view proxemics: distance and position sensitive interaction. In *Proceedings of the 2nd ACM International Symposium on Pervasive Displays*. 1–6.

[4] Kevin Fan, Jochen Huber, Suranga Nanayakkara, and Masahiko Inami. 2014. SpiderVision: extending the human field of view for augmented awareness. In *Proceedings of the 5th Augmented Human International Conference*. 1–8.

[5] Tian Gan, Yongkang Wong, Daqing Zhang, and Mohan S Kankanhalli. 2013. Temporal encoded F-formation system for social interaction detection. In *Proceedings of the 21st ACM international conference on Multimedia*. 937–946.

[6] Maia Garau, Mel Slater, Simon Bee, and Martina Angela Sasse. 2001. The Impact of Eye Gaze on Communication Using Humanoid Avatars. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Seattle, Washington, USA) *(CHI '01)*. Association for Computing Machinery, New York, NY, USA, 309–316. https://doi.org/10.1145/365024.365121

[7] Maia Garau, Mel Slater, Vinoba Vinayagamoorthy, Andrea Brogni, Anthony Steed, and M. Angela Sasse. 2003. The Impact of Avatar Realism and Eye Gaze Control on Perceived Quality of Communication in a Shared Immersive Virtual Environment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Ft. Lauderdale, Florida, USA) *(CHI '03)*. Association for Computing Machinery, New York, NY, USA, 529–536. https://doi.org/10.1145/642611.642703

[8] Jens Emil Grønbundefinedk, Christine Linding, Anders Kromann, Thomas Fly Hylddal Jensen, and Marianne Graves Petersen. 2019. Proxemics Play: Exploring the Interplay between Mobile Devices and Interiors. In *Companion Publication of the 2019 on Designing Interactive Systems Conference 2019 Companion* (San Diego, CA, USA) *(DIS '19 Companion)*. Association for Computing Machinery, New York, NY, USA, 177–181. https://doi.org/10.1145/3301019.3323886

[9] Uwe Gruenefeld, Tim Claudius Stratmann, Lars Prädel, and Wilko Heuten. 2018. MonoculAR: a radial light display to point towards out-of-view objects on augmented reality devices. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*. 16–22.

[10] Jan Gugenheimer, Evgeny Stemasov, Harpreet Sareen, and Enrico Rukzio. 2018. FaceDisplay: towards asymmetric multi-user interaction for nomadic virtual reality. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

[11] Edward T Hall. 1963. A system for the notation of proxemic behavior. *American anthropologist* 65, 5 (1963), 1003–1026.

[12] Hooman Hedayati, Daniel Szafir, and Sean Andrist. 2019. Recognizing F-formations in the open world. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 558–559.

[13] Hayley Hung and Ben Kröse. 2011. Detecting f-formations as dominant sets. In *Proceedings of the 13th international conference on multimodal interfaces*. 231–238.

[14] Shunichi Kasahara, Mitsuhito Ando, Kiyoshi Suganuma, and Jun Rekimoto. 2016. Parallel Eyes: Exploring Human Capability and Behaviors with Paralleled First Person View Sharing. In *ACM SIGGRAPH 2016 VR Village* (Anaheim, California) *(SIGGRAPH '16)*. Association for Computing Machinery, New York, NY, USA, Article 16, 2 pages. https://doi.org/10.1145/2929490.2929495

[15] Adam Kendon. 1990. *Conducting interaction: Patterns of behavior in focused encounters*. Vol. 7. CUP Archive.

[16] Feng Liang, Stevanus Kevin, Kai Kunze, and Yun Suen Pai. 2019. PanoFlex: Adaptive panoramic vision to accommodate 360° Field-of-view for humans. In *25th ACM Symposium on Virtual Reality Software and Technology*. 1–2.

[17] Christian Mai, Lukas Rambold, and Mohamed Khamis. 2017. TransparentHMD: revealing the HMD user's face to bystanders. In *Proceedings of the 16th International Conference on Mobile and Ubiquitous Multimedia*. 515–520.

[18] Nicolai Marquardt, Ken Hinckley, and Saul Greenberg. 2012. Cross-device interaction via micro-mobility and f-formations. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*. 13–22.

[19] Paul Marshall, Yvonne Rogers, and Nadia Pantidi. 2011. Using F-formations to analyse spatial patterns of interaction in physical environments. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*. 445–454.

[20] Kana Misawa and Jun Rekimoto. 2015. Chameleonmask: Embodied physical and social telepresence using human surrogates. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. 401–411.

[21] Florian Mueller, Sophie Stellmach, Saul Greenberg, Andreas Dippon, Susanne Boll, Jayden Garner, Rohit Khot, Amani Naseem, and David Altimira. 2014. Proxemics Play: Understanding Proxemics for Designing Digital Play Experiences. In *Proceedings of the 2014 Conference on Designing Interactive Systems* (Vancouver, BC, Canada) *(DIS '14)*. Association for Computing Machinery, New York, NY, USA, 533–542. https://doi.org/10.1145/2598510.2598532

[22] Jonathan Mumm and Bilge Mutlu. 2011. Human-Robot Proxemics: Physical and Psychological Distancing in Human-Robot Interaction. In *Proceedings of the 6th International Conference on Human-Robot Interaction* (Lausanne, Switzerland) *(HRI '11)*. Association for Computing Machinery, New York, NY, USA, 331–338. https://doi.org/10.1145/1957656.1957786

[23] Jason Orlosky, Qifan Wu, Kiyoshi Kiyokawa, Haruo Takemura, and Christian Nitschke. 2014. Fisheye vision: peripheral spatial compression for improved field of view in head mounted displays. In *Proceedings of the 2nd ACM symposium on Spatial user interaction*. 54–61.

[24] Yun Suen Pai, Benjamin I Outram, Benjamin Tag, Megumi Isogai, Daisuke Ochi, and Kai Kunze. 2017. GazeSphere: navigating 360-degree-video environments in VR using head rotation and eye gaze. In *ACM SIGGRAPH 2017 Posters*. 1–2.

[25] Magdalena Rychlowska, Leah Zinner, Serban C Musca, and Paula M Niedenthal. 2012. From the eye to the heart: eye contact triggers emotion simulation. In *Proceedings of the 4th Workshop on Eye Gaze in Intelligent Human Machine Interaction*. 1–7.

[26] Lichao Shen, MHD Yamen Saraji, Kai Kunze, and Kouta Minamizawa. 2018. Unconstrained neck: Omnidirectional observation from an extra robotic neck. In *Proceedings of the 9th Augmented Human International Conference*. 1–2.

[27] Dana Tokmurzina, Nurbolat Sagitzhan, Abzal Nurgaliyev, and Anara Sandygulova. 2018. Exploring Child-Robot Proxemics. In *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (Chicago, IL, USA) *(HRI '18)*. Association for Computing Machinery, New York, NY, USA, 257–258. https://doi.org/10.1145/3173386.3177083

[28] Pamela Whitten, Debra J Ford, Nancy Davis, Renee Speicher, and Bart Collins. 1998. Comparison of face-to-face versus interactive video continuing medical education delivery modalities. *Journal of Continuing Education in the Health Professions* 18, 2 (1998), 93–99.

[29] Jacob O Wobbrock, Leah Findlater, Darren Gergle, and James J Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 143–146.

[30] Katrin Wolf and Lars Lischke. 2014. Urban proxemics for public guidance. In *Proceedings of the NordiCHI Workshop on Ubicomp beyond Devices: People, Objects, Space and Meaning, ACM*, Vol. 10. 2639189–2654842.

[31] Dingding Zheng, Laura Lugaresi, George Chernyshov, Benjamin Tag, Masa Inakage, and Kai Kunze. 2017. Wearable aura: an interactive projection on personal space to enhance communication. In *Proceedings of the 2017 ACM UbiComp Adjunct*. 141–144.