EDITOR: Mike Potel, potel@wildcrest.com

DEPARTMENT: APPLICATIONS

The Making of a Newspaper Interview in Virtual Reality: Realistic Avatars, Philosophy, and Sushi

Ramon Oliva, *Event Lab, University of Barcelona, 08035, Barcelona, Spain*

Alejandro Beacco, *Universitat Politècnica de Catalunya, 08034, Barcelona, Spain*

Jaime Gallego and Raul Gallego Abellan, *Event Lab, University of Barcelona, 08035, Barcelona, Spain*

Mel Slater, *University of Barcelona, 08035, Barcelona, Spain*

VR United is a virtual reality application that we have developed to support multiple people simultaneously interacting in the same environment. Each person is represented with a virtual body that looks like themselves. Such immersive shared environments have existed and been the subject of research for the past 30 years. Here, we demonstrate how VR United meets criteria for successful interaction, where a journalist from the Financial Times *in London interviewed a professor in New York for two hours. The virtual location of the interview was a restaurant, in line with the series of interviews published as "Lunch with the FT." We show how the interview was successful, as a substitute for a physically present one. The article based on the interview was published in the* Financial Times *as normal for the series. We finally consider the future development of such systems, including some implications for immersive journalism.*

On 5 February 2022, Mr. John Thornhill, (Innovation Editor of the *Financial Times*, London, U.K.) interviewed Professor David Chalmers (Professor of Philosophy and Neuroscience, New York University, USA) for more than two hours. The interview was focused on the latter's book concerned with the nature of reality and virtual reality (David Chalmers, *Reality+: Virtual Worlds and the Problems of Philosophy*, 2022). The interview itself took place in virtual reality (VR), with the protagonists in London and New York, each represented by virtual human bodies that had a strong likeness to their real bodies. The interview was simultaneously attended by two others located in Barcelona, Spain, invisible, but recording video of the interview. In this article, we describe the system, VR United, that made this possible, the outcome of the interview, and relate this to other work on VR shared between remote participants.

VR as a shared experience with several remotely located people present together in the same virtual environment goes back to the wave of interest in VR in the late 1980s and early 1990s, "Reality Built for Two" supported multiple people meeting together in the same environment each represented by a virtual body, with the limitation on the number of people imposed by the complex hardware system setup.[1] Since then, there has been a massive amount of research both on the technical side of how to set up a multiperson system, and on its effects on participants.

There was a flowering of such systems in the 1990s, which were termed "Collaborative Virtual Environments" (CVE). Benford et al.[2] reviewed seven examples, ranging from a military oriented one through to the MASSIVE systems (1 and 2) that were oriented toward television and online entertainment.[3] The Distributed Interactive Virtual Environment (DIVE)[4]

Digital Object Identifier 10.1109/MCG.2023.3315761
Date of current version 7 November 2023.

system was particularly notable, used over several years by many groups particularly in European funded projects,[5] and several experimental studies were carried out to understand participant reactions to such virtual meeting spaces, for example Tromp et al.[6]

An early article about DIVE[7] introduced a number of requirements that any such VR system must address, still of importance today.

- *Simultaneous Presence*: The system must afford multiple people being simultaneously present in the same virtual environment perceiving the same situation and events, although each from their own unique embodied perspectives.
- *Embodiment*: Participants must be embodied in virtual bodies ("avatars") that at least indicate their locations to each other, otherwise spatial interaction between people is not possible—even though these representations might graphically be very simple, referred to as "blockies" in DIVE.
- *Objects*: Participants must be able to create, modify, and interact with objects.
- *Networking*: There must be a networking solution that maintains a consistent and persistent representation of events, so that the actions of participants can be perceived by others with minimal delay. It is essential that the network latency be at a low enough level so that the world consistency is maintained. Different systems have different network models, such as peer-to-peer communications, multicast solutions, centralized or decentralized, with different client–server policies. Rapid communication between peers is essential for interaction and collaboration in such environments.
- *Scalability*: The networking must be scalable, so as to not degrade with increasing participant numbers.
- *Collisions*: There must be collision detection to support both interaction with objects and between participants.
- *Communication*: There must be a method for participants to communicate with each other. Minimally this might be through changes in location and gestures, but more importantly with voice and the ability to share documents.
- *Locomotion*: Participants must be able to move through the virtual environment, with an understandable interface to permit this.

This article features a single case study detailing the experiences of a professional journalist with limited VR experience conducting a full two-hour interview on their own with an interviewee, to gather material for publication of an article in a major international newspaper. This application presented unique real-world challenges beyond those typical in an experimental study, which would have been limited to role-playing participants conducting short mock interviews without real-world objectives.

In the next section, we describe the technical features of VR United, both in general, and what had to be done for the creation of the scenario for the interview. Then we describe the interview itself. In the "Discussion" section, we return to the issue of the findings of previous studies, and how this particular application relates to those. We also relate the experience to the issue of "immersive journalism"—how the mass advent of VR might be reflected in journalistic practices and opportunities.

VR UNITED KEY FEATURES

Here we consider each of the requirements discussed in the introduction and show how the VR United system realizes those.

Simultaneous Presence

VR United is a platform that delivers shared virtual experiences, nowadays referred to as a Metaverse. The application can be downloaded from https://github.com/eventlab-projects/VRUnited_Builds with versions compatible with the most common head-mounted displays on the market (Meta Quest, Pico, Oculus Rift, and any OpenXR compatible device such as HTC VIVE or Valve Index). VR United can also be used on a regular computer with no VR support.

Furthermore, anyone can use the VR United SDK to add new scenarios and avatars to the official distribution or create a custom distribution. We have successfully used this application in a variety of remote application such as (but not limited to) virtual meetings and conferences, concerts, two people playing chess remotely, and journalistic interviews, as in this article.

Embodiment

A critical aspect is that the platform employs an accurate representation of the participants as virtual human characters that closely resemble them. In the particular application discussed here, the people involved were far away from our technicians (London and New York versus Barcelona), and we wanted to avoid any travelling at all as well as any cumbersome scanning technology. Therefore, we took a different and simpler approach requiring only a single input picture of each participant. Since nowadays everybody has a smartphone with a built-in camera, retrieving a

FIGURE 1. Photographs and derived full body avatars of the interviewee (left two) and the interviewer.

frontal picture of each participant with a specific pose was straightforward.

More specifically, we used an automatic system to obtain a realistic 3-D avatar reconstruction of a person using only the frontal RGB image.[8] This workflow first determines the pose, shape, and semantic information from the input image using state-of-the-art deep learning methods. This information is processed to create the skeleton and the 3-D skinned textured mesh that forms the final avatar. Additionally, a specific head reconstruction method is used to correctly match the final mesh to a realistic avatar. This way, we obtained each avatar in around 30 minutes (Figure 1).

The avatars were embodied by the interviewer and interviewee, so that when in the VR they looked down toward themselves they would see the virtual body substituting their real one. Moreover, the other person would, of course, look like the actual real person.

The participants both had access to a Meta Quest 2 head-mounted display. They were seated throughout, so that real-time body tracking using the Quest controllers tracked the hands, and together with real-time head tracking reconstructions of the head and upper body movements were obtained using inverse kinematics as described in Oliva et al.[9]

In general, when the VR United application starts, participants choose the avatar to represent themselves in the VR. Virtual mirrors can be introduced into the scenes, so participants can see their reflections. We provide a generic male and female avatar for different races (Caucasian, Afro-American, and Asian). However, participants can also generate their realistic look-alike avatar with the method previously described. In that case, they receive a unique code to unlock their personal avatar and so they can join the shared environment using their own look-alike virtual body. The system saves their selection, so the next time they run the application, that avatar will appear selected by default.

However the avatar is chosen, we use the QuickVR library for body tracking.[9] One of the main features of this library is that, as well as VR United, it is cross-platform and it adapts automatically to the available tracking data provided by the different VR systems being used by the participant. This means that we are not only limited to tracking the rotation and position of head and hands, but also can track a participant's feet, fingers, eyes, or facial expression if that data are provided by the VR devices.

Objects

When creating a scene for VR United, developers can define some objects with which participants can interact. The state of those objects (mainly position and rotation, but also other properties such as scale and color) are synchronized using the underlying networking engine.

For the sake of clarity, consider as an example the virtual sushi restaurant scenario used for this interview. There, the participants could select pieces of sushi that were on the virtual table at which they were seated. When participants grabbed a piece of sushi by moving their virtual hand over it and closing it with the controller buttons, they could move it through the space. When they moved it near the position of their mouth, it would disappear as if eaten. This is also replicated on the other clients, enhancing the feeling that all of them are sharing the same exact world.

Networking

We use Photon Network (https://www.photonengine.com) as the underlying network engine to manage client synchronization, a robust and well proven tool used by thousands of commercial applications. When using Photon, all client communication is performed through the servers provided by Photon. This greatly simplifies the setup as we need only ensure that the clients have an internet connection, so they can reach Photon servers, and that they are all connected to the same Photon server. All the networking low level details (firewall rules, reliability, load balancing...) are transparent to the programmer.

Once the connection to the server is completed, participants have to create a *Room* or join an existing one. A room can be considered as a specific instance of the application and the client that creates the room is considered to be the *Master Client*. In order to share the same virtual experience, participants must agree in advance about the room they must all join.

Scalability

Two factors determine the scalability of the application. First of all, Photon servers have a hard limit of concurrently connected users (CCU), which is from 20 for the free version up to 2000 for the most expensive plan. However, in all cases there is a maximum of 16 users per room. Another option is to host a copy of the Photon server. In that case, there is a limit of 100 CCU for the free plan and an unlimited number of CCU for the most expensive plan. When hosting a Photon server, the maximum number of participants per room is equal to the CCU of the plan. As VR United is under continued development, we are using the 20 CCU plan hosted by Photon, but we are studying other plans for later uses of the application.

Other factors that limit the scalability are the characteristics of the application and most importantly, the amount of data that has to be transmitted each frame for client synchronization. We have successfully tried VR United with up to ten participants sharing the same environment. We start to experience performance issues when there are 11 or more participants in the same room. However, this is without any special optimization, so the real limit would be much higher.

Collisions

In many scenarios of VR United, participants are seated with hands initially resting on a tabletop. However, the real table height may differ from participant to participant. In order to solve this problem, when the scenario is loaded a calibration process is executed which adjusts the height of the virtual table so its top is just below the hands of the avatar. During the calibration process, participants must have their hands resting on the top of the real table. This ensures a match between the virtual and the real table and artifacts such as intersecting the table or not reaching it due to height differences can be avoided. There are also scenarios in which participants can interact with objects, which have physical properties that determine how they collide with other objects in the scene, and also participants can move through the virtual environment and can physically interact with other participants and virtual objects. Collision detection and physics interaction is driven by Unity's physics engine (which is an integration of NVIDIA PhysX). In order to guarantee that all participants see exactly the same, the physics engine only acts on the client that owns the object (which by default, is the *Master Client)*, while the remaining clients simply set the position and rotation of their instance of that object that is sent by the owner. At the time of writing the Meta Quest 2 has an experimental feature that allows the user to define the spatial layout of their room including the desks, and Meta Horizon Workrooms specifically allows participants to define the height and extent of their real desk.

Communication

Participants can talk to each other as if they were sharing the same real space. Sound was spatialized, a factor found to be important for social presence,[10] and mitigated according to distance, using Unity3D settings. Photon Voice was used for sound synchronization (i.e., transmitting the sound (voice) from one client to the others as it integrates seamlessly with Photon Network). So only a microphone is needed (which is included in devices such as Meta Quest or Pico families).

As mentioned earlier, we use QuickVR for body tracking, which adapts to the data provided by the different VR devices and so we can track eye direction, blinking and even facial expressions if those devices support it. Otherwise, those features are simulated by software. These support nonverbal communication, which is crucial in order to enhance realism of the interactions. Speaking engages eye movements, head gaze direction, and gestures with the arms and body. All of these are supported by default; participants do not have to do anything special, only talk as normal. This is really very different from talking with people through online teleconferencing applications such as Zoom or Meet, where participants are represented by video windows showing only their face. Unlike the system presented here, these cannot easily show the bodily nonverbal communications that accompany speaking, and absolutely cannot involve spatial interaction—such as, for example, to which person gaze is directed, or spatial proximity between people.

Finally, participants have the possibility of sharing an external computer screen in VR United, making it possible for example to do presentations inside VR, so one could organize an international workshop completely online.

Locomotion

Although the application presented here was a seated experience, it is possible also to include locomotion in VR United scenarios, so participants can move through the virtual environment. For this feature, we also use QuickVR, which offers several locomotion solutions such as moving through the virtual environment with the real displacement of the participant, such as their walking, but also using different metaphors when the real place has limited dimensions, such as moving using the controllers, performing walking-in-place, or using a teleport ray.

RESULTS

The Conversation

Figure 2 shows images from the conversation captured by the two people in Barcelona who were filming the event. The pictures show that the conversation was animated, with participants naturally using hand gestures, and generally upper body gestures to emphasize their points, as in a normal conversation. This together with gaze direction based on the head tracking emphasizes the fundamental difference between the VR approach and, for example, online teleconferencing, where typically only the faces can be seen, and there is no sense of the spatial arrangement of the participants, nor their gaze directions.

Publication

An important sign of success is that the article derived from the conversation was published in the *Financial Times* (11 February 2022 by John Thornhill). This was as part of a series called "Lunch with the FT" and the length of the article was in line with others published in this series. Normally such interviews are carried out over lunch in a restaurant. Hence, we reproduced a restaurant setting, and at the request of the participants displayed sushi on the plates.

A further indication of success is that the interview continued without interruption (except for two small glitches of a few seconds) for two hours. The participants did not complain of any discomfort, and the conversation very quickly developed into a normal flow about Prof. Chalmers' views on reality and VR.

Presence

The participants commented about their experience, which they discussed near the start and the end of the session. This can be seen on video (https://www.youtube.com/watch?v=1dACicAYdYg). For example, there was a discussion on presence:

> JT: I was very interested by the sense of presence. Do you think you had presence in the VR?
> DC: . . .Totally, the sense of being actually here in this space, I think I have kind of got that now. It gets broken occasionally when my body twists into weird shapes or I find myself suddenly floating 6 feet in the air, that tends to break the feeling of presence, actually even at those moments, even when I was up in the air I still felt like—oh my gosh—here I am up there in the air so there is still something quite powerful to it, your body, your brain very naturally interprets these signals . . . as you being there in this space. How about you?
> JT: I know . . . I mean . . . on one level it all looks incredibly artificial, but on another level, you just feel as though there is a reality. I am sitting opposite you at a table, I can hear your voice perfectly, there is a sense that you are sitting there.
> DC: And the avatar looks actually quite realistic, maybe it helps that we have not met in the physical world. It is not unlike having lunch with somebody, at least until we start eating.

Toward the end of the two-hour session:

> DC: For the most part . . . it does feel like we are in a real space which is very intense, and you are right at the center of my attention. . .

FIGURE 2. Images from the conversation showing upper body engagement through gesturing. (A) The interviewee (DC) responding to the interviewer (JT). (B), (C) The interviewee gestures while talking. (D) Over the shoulder viewpoint from the interviewer.

Eating

The participants were able to pick up the virtual sushi using the controllers making a grasping gesture, to bring it toward their mouths, at which point it would disappear:

> DC: Very good I ate the sushi, now I feel like I ought to eat a sushi in the physical world . . . more or less synchronously . . . quite good! . . . if you had the hand tracking on there would be some hope you could actually do these two things synchronized with each other, but as it is I am virtually eating a virtual sushi followed five seconds later by physically eating a physical piece of sushi.

Later, there was some sign that normal politeness conventions were being adhered to:

> DC: There are two uneaten sushi plates here, the table was actually set for 4, I am occasionally tempted to reach out for one of those sushis, but it still feels like I am eating Mel's or Ramon's sushi. . .

Glitches

On two occasions the virtual body of DC became distorted. This occurred because he had to carry out an action in physical reality—once to close a door, and another time to plug in his HMD because it was running out of battery. On each occasion he moved out of the Quest guardian boundary, and the tracking was lost. On the first occasion, this was rectified by his leaving and rejoining the meeting, and on the second occasion it self-rectified.

DISCUSSION

We have described the use of the VR United system to enable a face-to-face interview of a professor by a journalist 5585 km (3470 miles) apart, both immersed in the same virtual restaurant. The scenario resulted in a newspaper article in the same way as if the protagonists had met physically. They commented favorably about their sense of presence in the environment and copresence with each other.

The conversation was carried out in a natural way with hand and upper body gestures, and it lasted for about two hours—a long time for two people to interact within VR. Body gestures in VR with people communicating via voice and their virtual bodies (i.e., bidirectional nonverbal communication accompanying the verbal) has been found to lead to greater verbal communication compared with static nonexpressive avatars.[11]

Social Presence

Oh et al.[10] carried out a systematic review of social presence (the feeling of being with a remote other) based on 152 studies, concentrating on the predictive factors. An important factor turned out to be visual representation coupled with a consistent behavioral realism. In other words, high-quality photorealistic representation of the other person combined with high-quality behavioral realism is more likely to lead to social presence than when combined with low-quality behavioral realism. Another important factor relevant to the present application was found to be virtual physical proximity between the protagonists. The greater the proximity, the greater the social presence. Each of these factors were present in our scenario—participants looked like their real selves, had full upper-body gestural capability, and were seated close to one another. With respect to gestural capability, it is important to note that this was not explicitly mentioned to the participants as a capability that they could exploit. They just carried out a conversation as normal, and their gestures were automatic. Hence, the gestural capability was implicit and transparent to the participants—they simply conversed without having to consciously exploit the fact that their upper body movements were mapped to their avatar, any more than we have to think about this in physical reality.

Gamelin et al.[12] compared two avatars in a collaborative task in VR. One was a 2.5-D point cloud integrated into the environment with respect to depth, but with very high fidelity motion tracking (essentially equivalent to video). The second was a standard 3-D avatar, which moved according to tracking, and thus based on a fixed number of tracking points. They found that the 2.5-D point cloud avatar with its greater motion fidelity resulted in improved collaborative task performance compared with the 3-D avatar. The importance of self-avatar representation that moves synchronously with the real movements of participants was also reported in two experiments by Dodds et al.[11]

The very fact that participants had a virtual body that moved synchronously and in correspondence with their movements was of intrinsic importance. It has been known for a long time that self-body representation enhances presence.[13] Moreover, Steed et al.[14] found that having a virtual body that moves according to the participant's movements enhances the cognitive processes of participants compared to not having a mapped body that reflects their movements. Participants were better able to recall pairs of letters in the active avatar condition compared to a passive one, or where they were instructed to move their hands or not. Such potential cognitive enhancement would have been advantageous in a situation where the interviewee had to comment in detail on complex philosophical issues regarding the nature of reality.

Having a virtual body also enhances trust between participants. A study by Pan and Steed[15] showed that a virtual body representation enabled participants to complete tasks together faster compared to a no-body representation, based on enhanced trust. Collingwoode-Williams et al.[16] further found that consistency of representation, where each participant was represented by an avatar, compared to at least one of them only represented by a representation of the hand-held controllers further enhanced trust.

Eating in VR

As DC mentioned during the interview, it would have been possible to combine the virtual and actual eating of his sushi. This could be enabled by direct hand tracking instead of using the controllers. Thus, when the participant would grab the sushi and lift it to his mouth, the same transformations would be applied to the sushi as to the hand. This would only work if the participants really had taken hold of the sushi, rather than only put their hands near the plate without actually grabbing the food. However, it would be possible to track the sushi itself using inside-out vision-based tracking, provided that the system had been pretrained for this. However, there would need to be tracking accuracy in order to place the virtual food correctly in the mouth in synchrony with the real food.

This strategy was effectively used by Oliver and Hollis,[17] who tracked the hands of participants using Leap Motion for the hand—including finger tracking—and also had the location of objects such as the table top, plates, and pieces of pizza using VIVE trackers. Hence, when participants lifted real food to the mouth this could be properly represented in the VR. Nevertheless, the authors noted that some participants did have problems in accurately placing the food in their mouths.

A quite different method was used by Korsgaard et al.[18] Whenever their system detected that participants were looking toward food on a plate, they switched the view seen through the HMD to actual video of the food. Whenever the viewpoint was directed away from the plate, then the VR scenario would be displayed instead. In this way, participants could simultaneously engage in VR while eating—although the visualization of the eating process was via video of physical reality rather than virtual reality. Although this addresses the problem of eating in VR, it does not maintain presence in the VR scenario. This approach could be more naturally realized using augmented reality, since the participants would see the real physical environment, and could just eat their food in a normal way. Moreover, each participant would see the remote person virtually sitting across the table, as if they were together in the same physical space. It is quite feasible that in the near future such an interview could have the journalist sitting in a sushi bar in London, and the interviewee in New York, and each interact with the other as if together in the same physical restaurant.

In addition to using VR to simulate the eating of food and actually eating it, there has been work on using VR to change the environment so as to enhance the dining experience.[17] Other approaches use VR or AR to change the flavor of the food—so that what might in reality be a very simple bland piece of food, through visual and olfactory augmentation the taste of the food can be transformed, as studied in "Project Nourished" http://www.projectnourished.com.

Although our scenario did not exploit synchronous real and virtual eating of the sushi, it is clear that this could be achieved.

Avatar Appearance and Representation

While many studies have looked at the impact of realism of the avatar,[19] the question of how much the avatar has to look like the self has not been investigated. Here the experience suggests that the likeness of the avatars to the participants may have been an important contributor to the success of the scenario. In an interview after the VR experience, JT pointed out the importance of the avatar representation: "One of my colleagues ... interviewed Nick Clegg of Facebook, or Meta now, in the metaverse using Facebook technology a few weeks ago, but they had avatars that were representing them. I think what was different with this was lifelike avatars, so I really did get the sense that David looked like David. I think it took a bit of a while to acclimatize to this new environment, but once we had done so it lent itself to a kind of intimate conversation."

In the near future very high-quality automatic 3-D capture of the face and facial expressions from depth cameras will be available as a matter of course (e.g., currently advertised for the Apple Vision Pro). The extent to which this makes any difference to the quality of resulting communication between people compared with simpler methods will be an important area of research.

We used the Meta Quest 2 head-mounted display, but already at the time of writing it is possible to go beyond what that offered in ways that could significantly improve remote immersive interpersonal communications. Since the interview scenario, Meta has greatly improved the hand tracking capability, so that the handheld controllers would not be necessary today. Moreover, Meta Quest Pro and Pico Enterprise have built in eye and facial expression tracking and these features are already incorporated into VR United. For remote face-to-face communication the only block on using these more advanced features is whether participants have access to these types of devices. Clearly though, these features are going to become the norm in the relatively near future.

Implications for Journalism

The idea and an example of immersive journalism was introduced by de la Peña et al.[20] Since then, this field has taken off and 360° videos are widely used by several media outlets. Immersive journalism has been concerned with how news stories are presented to the public (i.e., immersively). What we have shown here is that there is another side of this, which is how journalists might *make use of* shared VR in the very production of their stories. A further consequence is that a scenario captured in VR can then be experienced immersively by the public.

Using this type of technology, journalists and media outlets could, with a VR technical support team, have no limitations to interview people in any location of the world that normally would have difficult, limited, dangerous, or restrictive access. Journalists, as the results of the scenario indicate, could experience interviewing people almost feeling like a real face-to-face meeting. However, VR could create some limitations for journalists to pick up subtleties such as small but meaningful changes in facial expressions and eye movements, the nervousness of the interviewee, etc. Nevertheless, the fact that the interview is in VR with avatars could as well break certain boundaries that maybe being face-to-face the interviewer would not dare to cross: the very fact of not being face-to-face in reality might allow interviewees to deliver information that they would be reluctant to give away face-to-face. This is very similar to the fact that patients with a phobia (e.g., of speaking in public) will nevertheless speak to a virtual audience even though this generates similar anxiety as talking to a real one. Here the perceived gap between reality and virtual reality might be an advantage.

This technology would also bring new possibilities for the audience. They could simply watch a recorded (or live) video of the interview as normal or immersively as a 360° version. However, members of the public could also be virtually present in a live VR interview (e.g., as with interviews that take place in the street, with crowds watching and listening), or join a VR replay of the interview. Audience members could also navigate through the environment, leave recorded messages for others to witness later, and so on. This would maintain the interview as a live event, since each later participant could see not just the interview but embodied audience reactions. Here the interview would be a persistent virtual environment, something that would be almost impossible to do with playbacks of 360 video, since each intervention would require the production of a new one.

CONCLUSION

There are several avenues for further development of VR United in this domain. First, while there was lip sync, facial expressions otherwise did not change. This will be rectified in the near future with the coming on stream of new HMD hardware that is likely to have built-in facial tracking. Second, we did not include eye tracking. Although this is supported by our Pico NEO devices, these were not available to the participants, who used the Quest 2.

It would be important to experimentally test the importance of the look-alike body representation, compared to arbitrary avatars of the same fidelity. Here there could be surprising results—possibly interviewees might be more open if they are not represented as themselves.

A third and vitally important area is that of security. Such immersive communication systems must be set up in a way that guarantees that identities cannot be hacked, that each participant in a discussion can be confident that they are really interacting with the person portrayed.

Overall, the interview of DC by JT for a newspaper article, as a virtual face-to-face meeting, opens up a new area of application of shared VR, not least contributing to a new form of immersive journalism.

ACKNOWLEDGMENTS

The author would like to thank Mr. John Thornhill and Professor David Chalmers for taking part in this application. VR United was developed under the European Research Council project "Moments in Time in Immersive Virtual Environments" (MoTIVE) (#742989). The particular application for the interview was funded under the Horizon 2020 programme H2020-FET-PROACT-2020-2 (#101017884) GuestXR.

REFERENCES

1. C. Blanchard et al., "Reality built for two: A virtual reality tool," *ACM SIGGRAPH Comput. Graphics*, vol. 24, no. 2, pp. 35–36, 1990.

2. S. Benford et al., "Collaborative virtual environments," *Commun. ACM*, vol. 44, pp. 79–85, 2001.
3. C. Greenhalgh, *Large Scale Collaborative Virtual Environments*. Berlin, Germany: Springer, 1999.
4. E. Frécon and M. Stenius, "DIVE: A scaleable network architecture for distributed virtual environments," *Distrib. Syst. Eng.*, vol. 5, no. 3, 1998, Art. no. 91.
5. V. Normand et al., "The COVEN project: Exploring applicative, technical and usage dimensions of collaborative virtual environments," *Presence—Teleoperators Virtual Environ.*, vol. 8, pp. 218–236, 1999.
6. J. Tromp, A. Bullock, A. Steed, A. Sadagic, M. Slater, and E. Frecon, "Small group behaviour experiments in the coven project," *IEEE Comput. Graphics Appl.*, vol. 18, no. 6, pp. 53–63, Nov./Dec. 1998.
7. O. Hagsand, "Interactive multiuser VEs in the DIVE system," *IEEE Multimedia*, vol. 3, no. 1, pp. 30–39, Spring 1996.
8. A. Beacco, J. Gallego, and M. Slater, "Automatic 3D avatar generation from a single RBG frontal image," in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces Abstr. Workshops*, 2022, pp. 764–765.
9. R. Oliva et al., "QuickVR: A standard library for virtual embodiment in unity," *Front. Virtual Reality*, vol. 3, 2022, Art. no. 937191.
10. C. S. Oh, J. N. Bailenson, and G. F. Welch, "A systematic review of social presence: Definition, antecedents, and implications," *Front. Robot. AI*, vol. 5, 2018, Art. no. 114.
11. T. J. Dodds, B. J. Mohler, and H. H. Bülthoff, "Talk to the virtual hands: Self-animated avatars improve communication in head-mounted display virtual environments," *PLoS One*, vol. 6, no. 10, 2011, Art. no. e25759.
12. G. Gamelin et al., "Point-cloud avatars to improve spatial communication in immersive collaborative virtual environments," *Pers. Ubiquitous Comput.*, vol. 25, no. 3, pp. 467–484, 2021.
13. M. Slater, B. Spanlang, and D. Corominas, "Simulating virtual environments within virtual environments as the basis for a psychophysics of presence," *ACM Trans. Graphics*, vol. 29, 2010, Art. no. 92.
14. A. Steed, Y. Pan, F. Zisch, and W. Steptoe, "The impact of a self-avatar on cognitive load in immersive virtual reality," in *Proc. IEEE Virtual Reality*, 2016, pp. 67–76.
15. Y. Pan and A. Steed, "The impact of self-avatars on trust and collaboration in shared virtual environments," *PLoS One*, vol. 12, no. 12, 2017, Art. no. e0189078.
16. T. Collingwoode-Williams et al., "The impact of self-representation and consistency in collaborative virtual environments," *Front. Virtual Reality*, vol. 2, 2021, Art. no. 45.
17. J. H. Oliver and J. H. Hollis, "Virtual reality as a tool to study the influence of the eating environment on eating behavior: A feasibility study," *Foods*, vol. 10, no. 1, 2021, Art. no. 89.
18. D. Korsgaard, N. C. Nilsson, and T. Bjørner, "Immersive eating: Evaluating the use of head-mounted displays for mixed reality meal sessions," in *Proc. IEEE 3rd Workshop Everyday Virtual Reality*, 2017, pp. 1–4.
19. K. Zibrek, S. Martin, and R. McDonnell, "Is photorealism important for perception of expressive virtual humans in virtual reality?," *ACM Trans. Appl. Percep.*, vol. 16, no. 3, pp. 1–19, 2019.
20. N. de la Peña et al., "Immersive journalism: Immersive virtual reality for the first-person experience of news," *Presence, Teleoperators Virtual Environ.*, vol. 19, no. 4, pp. 291–301, 2010.

RAMON OLIVA is currently a postdoctoral researcher with the Event Lab, Universitat de Barcelona, 08035, Barcelona, Spain. He received his Ph.D. degree in computer science from Universitat Politècnica de Catalunya. Contact him at ramon.oliva@ub.edu.

ALEJANDRO BEACCO is currently an associate professor with the Universitat Politècnica de Catalunya, 08034, Barcelona, Spain. He was a postdoctoral researcher in the Event Lab, Universitat de Barcelona, Spain, at the time of carrying out this work. He received his Ph.D. degree in computer science from the Universitat Politècnica de Catalunya. Contact him at abeacco@cs.upc.edu.

JAIME GALLEGO is currently a postdoctoral researcher in the Event Lab, Universitat de Barcelona in Spain, 08035, Barcelona, Spain. He received his Ph.D. degree in image processing from Universitat Politècnica de Catalunya. Contact him at jgallego@ub.edu.

RAUL GALLEGO ABELLAN is a video journalist and currently a part-time Ph.D. student in the Event Lab at the Universitat de Barcelona, 08035, Barcelona, Spain. He received his B.Sc. degree in communication and media studies from Universitat Ramon Llull. Contact him at raulgaab@gmail.com.

MEL SLATER is currently a distinguished investigator in the Event Lab, Universitat de Barcelona, 08035, Barcelona, Spain. He received his D.Sc. degree in computer science from the University of London. He is the corresponding author of this article. Contact him at melslater@ub.edu.

Contact department editor Mike Potel at potel@wildcrest.com.