# BubbleVideo: Supporting Small Group Interactions in Online Conferences

Bill Rogers, Mark Apperley, Masood Masoodian

# BubbleVideo: Supporting Small Group Interactions in Online Conferences

Bill Rogers[1], Mark Apperley[1], and Masood Masoodian[2]

[1] Department of Computer Science, University of Waikato, Hamilton, New Zealand
{william.rogers,mark.apperley}@waikato.ac.nz
[2] School of Arts, Design and Architecture, Aalto University, Espoo, Finland
masood.masoodian@aalto.fi

**Abstract.** Increasing use of online conferencing systems, particularly over the past year, has highlighted problems in these systems, especially their poor support for small group interactions within larger meetings. These include clumsy small group formation (e.g., issues around joining and leaving existing groups), the difficulty of getting the correct level of audio isolation between groups, poor provision for shared editing of documents, as well as fatiguing aspects of video conferencing caused by presentation format and the necessity of remaining on camera view. This paper describes the motivation, design and implementation of a prototype online conferencing system, called *BubbleVideo*. Building on both virtual world and pure video paradigms, it implements an extensive 2D world with shared documents, in which users appear through real-time video, presented in "bubbles" that can be moved around. Users are given the possibility of deciding whether to join a group by viewing a conversation "leakage", which group members can share with outsiders.

**Keywords:** Online conferencing · video conferencing · group communication · group interaction · shared workspaces · small groups.

## 1 Introduction

Online conferencing is of great and increasing importance in many social, business and educational settings. Although growth in the use of online conferencing has been continuous over the past three decades, in more recent years there have been two waves of acceleration in its use: the global economic turn-down of 2008, and the COVID-19 pandemic starting in 2020. The first of these was driven by the need to reduce costs – particularly travel costs – and the second by the need to avoid physical group gatherings. In 2008, the poster child of online conferencing was Second Life [8], a massively multi-user computer game-like system in which people were represented by avatars – computer animated figures – that could be moved around in large simulated 3D worlds (see Figure 1, left).

Between 2008 and 2020, improvements in the performance of computer systems and internet meant that real-time video communication became possible for meetings involving 10's and even 100's of participants. One such conference

**Fig. 1.** Two alternative approaches to online conferencing, using avatars in Second Life (left) and video streams in Zoom (right).

system which provided support for a large number of participants is Zoom [19], which since 2020 has emerged as the most widely used platform for group meetings [9]. As such, the most commonly seen image of online conferences these days is with Zoom in its "Gallery" view, in which the screen is divided into a grid of separate video images – a grid of video talking heads (see Figure 1, right).

In this paper we present a review of several current systems, chosen to show the range of variations and combinations of the above mentioned two approaches that have been developed. In addition to basic online conferencing features, we pay particular attention to support for smaller sub-group interactions, and identify some communication and interaction elements which are not yet supported by most existing systems. We then present a prototype conferencing tool, called *BubbleVideo*, which we have developed to support some of these elements.

## 2 Alternative approaches to online conferencing

Online conferencing systems which aim to host a large group of participants need to not only provide the necessary communication channels – text chat, audio, and video – between the attendees, but also provide an environment in which the interactions between them takes place. Earlier systems, such as Second Life, provided minimum text chat and/or audio communication, in addition to a 2D or 3D space in which the participants could meet and share documents and workspaces. With the increase in speed and availability of high-bandwidth internet, real-time video communication has in more recent years been provided as a solution for all needs, while other interaction needs have largely been ignored [**?**,**?**]. Rather than seeing video communication as the ultimate solution, we would argue that both approaches have some clear benefits as well as disadvantages.

The main advantage of the Second Life-style conferencing systems was that their users were presented in a large 3D space – called the world – in which they could navigate – providing a "move to join" paradigm for choosing the groups in which they wished to participate. This 3D "world" modelled conventional physical conference meeting spaces, like conference presentation and meeting rooms. Detailed avatar models of participants were shown walking or seated. Avatar appearance could be roughly personalised, but their movement was generic. The

ability to fully participate in meetings was limited by the clumsiness of avatar control and the lack of feedback from facial expressions and other non-verbal communication forms that are typical of real-life [3].

A review by Erikson et al. [5] of a large online conference conducted in Second Life showed that whilst the conference was generally successful, and many parts worked well, there were also aspects of the experience that participants found less than satisfactory. For instance, the avatars were not accurate images of the people they represented, and although this can be liberating in a social context – allowing people to experiment with their appearance – it was not helpful in a business context. In this case, attendees would have preferred a better match between their real and avatar appearance. Another issue was that although keynote presentations worked in the 3D world setting, they would have been better as simple video presentations, because sometimes conference attendees accidentally intruded on the stage during presentations.

Perhaps the most common problem was, however, related to social interactions. While such interactions worked when they were well organised or had a focus (e.g., for poster sessions), more open social situations where people formed ad-hoc small groups were less successful. The scale on which distance-based audio isolation worked was problematic, with the wide-scale that was used leading to passers-by feeling like eavesdroppers, and participants feeling that their sound carried too far and could disturb others. Finally, it was noted that moving about with a fellow participant was not well supported in Second Life.

Zoom-style video systems, on the other hand, work best for single-group conferences. The use of video provides clear identification of participants and permits considerable non-verbal communication, albeit mostly limited to facial expressions and head movements. In comparison to Second Life-style conferences, however, the ease with which people can form, join and leave sub-groupings is limited, and it is difficult or impossible for a participant to find out what is happening in groups other than the one in which they are taking part (e.g., in a breakout room). The inability to quickly form ad-hoc groups can, for instance, severely limit social interaction outside formal conference sessions [14].

The large number of Zoom meetings and conferences since 2020 have provided opportunities for study of personal experiences of video conference sessions [11, 18]. For instance, Bailenson [1] hypothesises four non-verbal reasons for the phenomenon of "Zoom Fatigue", to try to explain why people find continuous long Zoom video conferencing sessions tiring. His reasons are:

- *"Excessive amounts of close-up eye contact is highly intense."* A group of face images filling a screen, all seemingly staring at you "simulates a personal space that you normally experience when you're with somebody intimately".
- *"Seeing yourself during video chats constantly in real-time is fatiguing."* Bailensen likens the experience to constantly having a mirror showing your reflection, which can make people more self-conscious and self-critical.
- *"Video chats dramatically reduce our usual mobility."* In a social gathering people are able to stand and move around. Even in a meeting room people are free to look around, work with documents on the table, or talk to others.

– *"The cognitive load is much higher in video chats."* Bailenson refers to the changed nature of non-verbal communication, the need to be aware of "staying on camera", and limitations in the way we can gesture.

With growing popularity, particularly since the start of the current COVID-19 pandemic [10], there are now a large number of online conferencing systems available. Table 1 provides a summary of some of the systems we have examined, chosen to illustrate the variety of interaction features provided, particularly as related to small group interaction. Our review involved both reading documentation and in all cases directly experimenting with the systems.

The systems reviewed here vary in the way in which sub-groups can be formed, and in the degree to which groups' audio and video streams are isolated from each other. In those that support avatar navigation, entering and leaving groups can be divided into proximity-based or pre-defined location-based groups. The use of proximity offers the option of users moving near a group in order to find out what is being discussed by them, without actually joining the group – i.e., to receive some "leaked" content. Most proximity-based systems allow group members to see other people approaching their group – although full 3D modelling as occurs in Second Life and Virbela leaves open the possibility of being "sneaked up on from behind". Systems with location-based audio isolation require users to join before learning anything about a group's conversation. None of the systems reviewed here appeared to be good at providing information about group conversations to outsiders, or in providing an easy way of insiders inviting outsiders into a group.

Documents can be shared in many of the systems we reviewed. The most common method is through screen sharing, in which the shares are either shown as part of the environment (e.g., on virtual screens), or as additional video streams. None of the systems appears to strongly support collaboratively editable documents (e.g., as shared workspaces). However, some systems such as Zoom permit remote users to control shared screens, or provide whiteboard-like tools editable by all group members.

## 3 Requirements for small group interaction

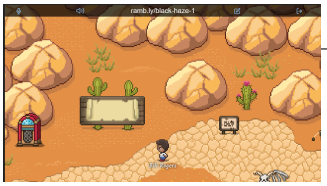
Our requirement gathering for an online conferencing system that would support small group communication and interaction started with an earlier prototype aimed at providing a 2D space for social gatherings during online conferences [14]. In addition, our experience of teaching online using existing conferencing systems during 2020-21 led us to widen the scope of its potential uses to include provisions for smaller group communication and interaction in online teaching, as well as other settings such as conferences and meetings. In particular, we wanted to support not only larger group communication and interaction (e.g., during a classroom lecture or a conference presentation), but also during smaller group activities such as a conference workshop or in-class group sessions.

Our experience of conducting such small group workshop-type activities using systems such as Zoom and Discord found them difficult or less than satisfactory

**Table 1.** Summary of the reviewed online conferencing systems.

**Virbela** [17]: Users navigate personalised avatars in a detailed 3D space. Audio is localised in pre-arranged spaces (e.g., theatres and poster areas).

An education system similar to Second Life. Groups defined by proximity, or pre-defined rooms/areas with audio isolation. Entry and exit is easy. Shared video and documents in presentation mode on simulated screens.

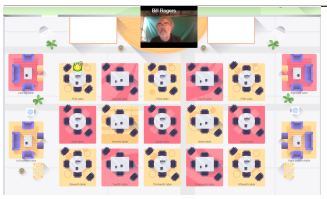




**Rambly** [12]: A 2D top-down world with person-like sprites used as avatars. Audio only. Supports proximity-limited audio.

Groups are proximity-based and it is easy to enter or leave them. No provision for document sharing.

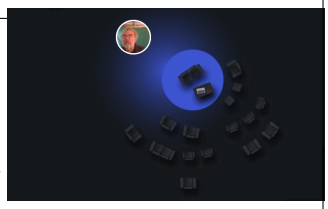
**Cozy Room** [2]: A 2.5D world of connected spaces (shown as orange border in the image) using icons (yellow ball) as avatars. Audio only. Audio isolated in each space.

Groups easy to enter and exit. Real-time creation and decoration of spaces is supported. Avatars have eyes that move to follow the mouse, giving a potential to invite bystanders into a group by looking at them. No provision for document sharing.



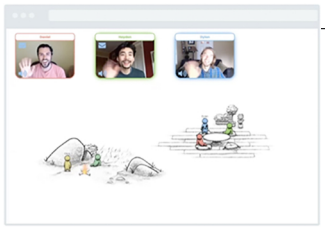
**Remo** [13]: Groups and navigable space are represented as an array of 'tables'. Users join a table to share audio and video, isolated from other tables.

Combined video conferencing and avatar navigation. Groups pre-defined, with easy navigation in and out. Screen sharing is possible using video streams(video displays can be enlarged).



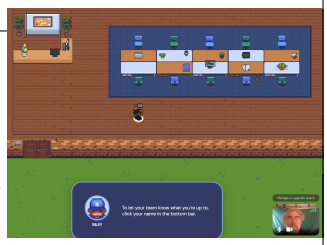
**Spacial Chat** [15]: The user's video image is used like an avatar for navigation in a single unified interface. Pre-arranged group areas with isolated audio communication.

Easy navigation in and out of group areas. Screen sharing is possible on in-world "screens".



**Topia** [16]: A sketch-style world providing focus areas for proximity-limited audio and video meetings. Abstract avatars are used for navigation.

Groups are proximity-based and limited in size (max 9). Navigation in and out of groups is easy. Although video and audio is limited to group members, it is possible to see the avatars of other groups that are nearby. Screen sharing is possible using video streams.

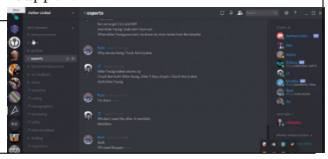


**Gather** [6]: A 2D top-down navigable world in which users have sprite avatars. Proximity limited audio and video windows along the screen top.

Groups formed by proximity. Easy to navigate in and out. Screen sharing allows access to documents, presented as enlarged video streams.

**VMX** [3]: A 3D meeting room, in which each user's video is projected on the face of an avatar animated by users body movement in real time (heads are see-through from behind).

Only one small group is supported but the system is interesting because of strong gesture and body position support.



**Discord** [4]: It was initially an audio and chat-based system structured in channels, which has now been extended to include video.

Depending on permission settings, any user can set up channels to support groups. Moving between channels is quick and easy, by clicking on a channel name in the list. Screen sharing is possible using video streams.

to use. For instance, the person in charge of a Zoom session has the responsibility of creating groups, and usually also assigning members to them. In Discord, channels are much better for this purpose, but still need to be created. Once the groups are formed in either of these systems, there is no way of maintaining an overview of their progress, other than visiting each one individually – which is slow to do in Zoom and somewhat intrusive in Discord.

Based on our own experiences and the review of existing systems, we defined the following requirements for a new prototype virtual gathering space which would support small group interactions and communication within larger gatherings. For the current prototype we decided to omit any provision of "host" control, and giving group participants equal rights. However, this might not be ideal in all situations (e.g., a teacher may need some controls in a class situation).

1. Use of real-time video rather than avatars to represent participants.
2. Support the "move to join" group formation paradigm, where a large space is provided and participants are free to form, join and leave groups.
3. Constrain audio within groups for private conversations, but also allow for overriding public audio across all groups for announcements or presentation.
4. Allowing side conversations in a setting where there is a principal speaker, without losing the main conversation.
5. Being able to share pre-prepared documents or live documents external to the conferencing system.
6. Provide support for collaborative editing of shared documents.
7. Allow people to move together in the conference space while maintaining a private conversation.
8. Provide a shared view of all the groups, so that each participant can see how people are grouped.
9. Help participants to know when it might be interesting or appropriate to join an existing group conversation.
10. Make it possible for people in a conversation group to be aware of others near by, but outside, their group who might be invited to join them.

## 4 *BubbleVideo* prototype

The meeting paradigm implemented in *BubbleVideo* is that of a conversation pit – i.e., like a large floor space. By walking around this space (using arrow keys, like a video game) a user may find (or create) a small group pit – an area to gather around and have a shared audio/video conversation, and perhaps share documents with everyone in the group to read or edit – as shown in Figure 2. By providing a single large space, *BubbleVideo* allows freedom of navigation over that space. Rather than moving avatars, users move their own small real-time video streams, which appear inside small "bubbles" on the floor space. This has the advantage of not separating movement from video conferencing – a single unified navigation/communication experience. Video bubbles are optionally tagged with users' names. The system provides for "zooming in" to get reasonably large video bubbles for more focused personal interaction – but the default

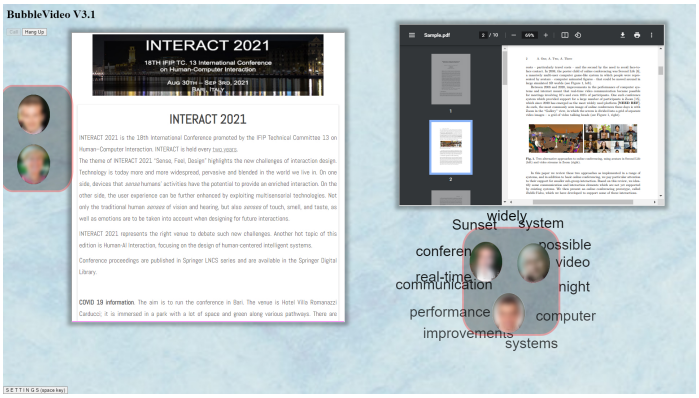



**Fig. 2.** *BubbleVideo* with two groups of participants (top left and bottom right), each sharing a document. The bottom right group is "leaking" their conversation.

is small images (110 by 90 pixels). To compensate for the small size, the system automatically centralizes and maximizes users' faces in their own video bubble, regardless of the actual wider video camera feeds. This improves the likelihood of faces being visible without users having to constantly try to keep on camera, giving them less reason to look at their own video. Video bubbles are also part of a larger screen space, where each user's location as well as other content such as shared documents are important. We hope that these factors will reduce the impact of the "array of staring faces", and the "intense eye contact" which both contribute to fatigue.

While audio conversations are by default global – initially all users receive audio from all other participants – users can however form audio bubbles in which their conversation can be fully or partially private. Both ad-hoc and round table-style audio bubbles are supported. In the ad-hoc form, an audio bubble is created when one user bumps their video bubble into another's and pauses for a moment. Once an audio bubble is formed by two users (grey areas around groups in Figure 2), others can join by simply moving into the group's bubble. Moving video bubbles away from each other immediately removes the group's shared audio bubble. Round table-style audio bubbles can be added from the conference settings dialog, and remain displayed on the floor space until explicitly removed. Users can navigate into table-style audio bubbles, as with ad-hoc bubbles. All audio bubbles resize automatically to accommodate their users as they join or leave. In addition, table-style bubble repel each other to maintain some separation. Users' positions in an audio bubble gently animate to avoid overlap. Users in an audio bubble can converse with each other, and also hear audio from outside their bubble. This makes it possible to establish a small private conversation between people who might also be listening to a (global) speaker from outside, if there is someone talking in the bigger space.

Although audio bubbles have the option of full privacy, in some situations such as a social gathering (e.g., a cocktail party [14]) small group conversations

do not necessarily need to be private, and are often open for outsiders to join. As such, a group may want people nearby to overhear all or, more usually, part of what is being said. Others may take advantage of overhearing parts of conversations to help them decide whether to join a group or not. *BubbleVideo* also allows group member to see who is nearby when they are inside an audio bubble. We hypothesize that small video bubbles in an active context will lead people to behave more like normal conversations in real-world – keeping eye contact by looking at video bubbles for some of the time, and gazing around at other times – reducing fatigue and also improving the likelihood of passers-by being noticed and maybe invited to join. Passers-by might in turn indicate an interest in joining by standing nearby and looking hopeful.

To support this form of social interaction, *BubbleVideo* includes a mechanism to allow aspects of a conversation to be "overheard" outside an audio bubble. When enabled by group members, this continuously runs automatic speech recognition on participants' audio streams, and extracts a randomized cloud of significant words to display around the group's audio bubble (see Figure 2 bottom right). Despite some speech recognition errors, the system "leaks" a sense of conversation topic, but only a general sense. The number of words tends to be quite high and they are presented in random arrangement, so it is usually difficult to see what exactly was being said. Our hypothesis is that the system hits an acceptable balance between privacy and useful "leakage".

Finally, *BubbleVideo* provides two forms of group shared workshpaces: screen shares and document shares, which can be presented in the navigation space (see Figure 2). Participants can move their video bubbles close to a shared workspace window – and form their group audio bubble next to it – or position their video bubbles elsewhere. A screen share can be used as a presentation, in which case, viewers can gather nearby and listen to the presenter. Document share, on the other hand, enables concurrent editing by group participants. The current implementation supports direct embedding of Google Docs [7] in the navigation space, to allow group participants to co-edit shared documents. This makes it possible to have all the shared editing and viewing options provided by Google Docs, supported by group interactions inside audio bubbles formed around them.

## 5 Conclusions

This paper presented *BubbleVideo* online conferencing prototype. Although we have used the system ourselves for discussions formulating this co-authored paper, our only other evaluation has thus far been limited to some informal tests of dynamic grouping with 14 participants. Overall it has functioned well and supported its initial design requirements. The remaining research questions are, how would participants make use of the grouping mechanisms, and how well does the leakage system work in providing conversation information and not violating privacy? A workshop style classroom trial is planned to test grouping, and conversation leakage will be evaluated using a word guessing game. The questions of supporting dynamic ad-hoc conversations will be tested in a "party" gathering.

## References


1. Bailenson, J.N.: Nonverbal overload: A theoretical argument for the causes of zoom fatigue. Technology, Mind, and Behavior **2**(1) (2021), https://doi.org/10.1037/tmb0000030
2. CozyRoom: CozyRoom (2021), https://cozyroom.xyz, (accessed in June 2021).
3. Dean, J., Apperley, M., Rogers, B.: Refining personal and social presence in virtual meetings. In: Proceedings of the Fifteenth Australasian User Interface Conference – Volume 150. pp. 67–75. AUIC '14, Australian Computer Society, Inc. (2014)
4. Discord: Discord (2021), https://discord.com/, (accessed in June 2021).
5. Erickson, T., Shami, N.S., Kellogg, W.A., Levine, D.W.: Synchronous interaction among hundreds: An evaluation of a conference in an avatar-based virtual environment. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 503—512. CHI '11, Association for Computing Machinery, New York, NY, USA (2011), https://doi.org/10.1145/1978942.1979013
6. Gather Presence Inc.: Gather (2021), https://gather.town/, (accessed in June 2021).
7. Google: Google Docs (2021), https://www.google.com/docs/about/, (accessed in June 2021).
8. Linden Research, Inc.: Second Life (2021), https://secondlife.com/, (accessed in June 2021).
9. Molla, R.: The pandemic was great for Zoom. What happens when there's a vaccine? (2020), https://www.vox.com/recode/21726260/zoom-microsoft-teams-video-conferencing-post-pandemic-coronavirus, (accessed in June 2021).
10. Oeppen, R., Shaw, G., Brennan, P.: Human factors recognition at virtual meetings and video conferencing: how to get the best performance from yourself and others. British Journal of Oral and Maxillofacial Surgery **58**, 643–646 (05 2020), https://doi.org/10.1016/j.bjoms.2020.04.046
11. Ramachandran, V.: Stanford researchers identify four causes for 'Zoom fatigue' and their simple fixes (2021), https://news.stanford.edu/2021/02/23/four-causes-zoom-fatigue-solutions/, (accessed in June 2021).
12. Rambly: Rambly (2021), https://rambly.app/, (accessed in June 2021).
13. Remo.co Virtual Events and Office space: Remo (2021), https://remo.co/, (accessed in June 2021).
14. Rogers, B., Masoodian, M., Apperley, M.: A virtual cocktail party: Supporting informal social interactions in a virtual conference. In: Proceedings of the 2018 International Conference on Advanced Visual Interfaces. AVI '18, Association for Computing Machinery, New York, NY, USA (2018), https://doi.org/10.1145/3206505.3206569
15. SpatialChat Ltd.: SpatialChat (2021), https://spatial.chat/, (accessed in June 2021).
16. Topia: Topia (2021), https://topia.io/, (accessed in June 2021).
17. Virbela: Virbela (2021), https://www.virbela.com/, (accessed in June 2021).
18. WIRED: A mission to make virtual parties actually fun (2021), https://www.wired.com/story/zoom-parties-proximity-chat/, (accessed in June 2021).
19. Zoom Video Communications, Inc.: Zoom (2021), https://zoom.us/, (accessed in June 2021).