Spatial Sound Concepts for F-Formations in Social VR

SYLVIA ROTHE, ROBIN WELSCH, and SVEN MAYER, LMU Munich, Germany

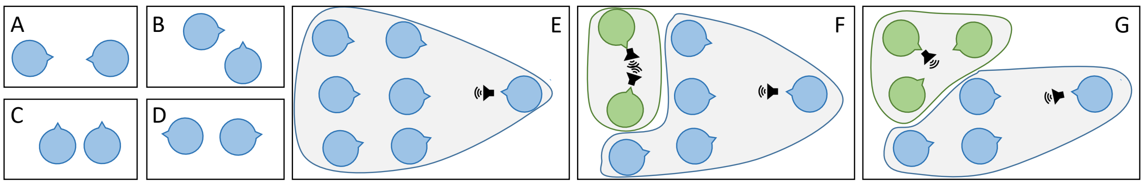

Fig. 1. F-formations of two people: (A) face-to-face conversation, (B) l-shape, e.g., interacting together on an object, (C) side-by-side, e.g., watching a movie, (D) back-to-back. Examples for sound zones: (E) lecture, (F) lecture and dialog, (G) lecture and discussion. Different sound zones are depicted by the colored-fill of silhouettes (green vs. blue).

Directional audio is key for fluent conversations in the virtual world. To allow directional audio, F-formations (facing formations) are core to understand the constellation between conversation partners. We present our approach to developing a sound concept in Social VR based on F-formations. For this, we introduce several F-formations and explain requirements for the sound design. We discuss our first experiences in observing several communication situations.

CCS Concepts: • **Human-centered computing** → **Virtual reality**; **Collaborative and social computing**.

Additional Key Words and Phrases: Virtual Reality, Social, Spatial Audio, Sound

1 INTRODUCTION

Currently, the main focus of virtual environments is on the visual component while the auditory component is unexplored [7]. While spatial audio has been explored in virtual reality (VR) [3, 13], we argue that spatial audio is especially important for Social VR environments. Here, different sound settings are applicable to communicate with others, e.g., a megaphone that can be heard by everyone, or loudness which decrease with distance. However, in Social VR there are a lot of communication situations where these settings are not sufficient. Besides the distances, the relative orientation and thus physical arrangement of humans to each other is important [9]. Such facing formations, called F-formations, are subject of sociological research and are influenced by the task that people want to do together (e.g., competitive, collaborative, or communicative) [6, 9, 12]. Figure 1 (A-D) shows four examples of such F-formations for two people; however, their voices can be a problem for others [4]. Key in this is that on nonverbal behavior aligns with intimacy [14] of the social interaction. This is typically done by means of distance [5], orientation [10] and gaze [11]. Interestingly, gaze produces higher levels of intimacy and therefore [1], which has also shown to persist in virtual environments [2]. How such psychological variables can be integrated into the soundscape of social VR has yet not been tested. Thus, sound concepts are needed that can facilitate social interaction but also suit the social setting.

2 COMMUNICATION SITUATIONS

We identified several typical situations that are possible in Social VR, e.g., (1) lecture: one person is speaking to a group of people, (2) dialog: two people are speaking to each other, (3) discussion: three or more people are speaking with each other. Each of these situations need their own sound design: In situation (1), one person should be audible for all other people. In most cases, a dialogue (2) should only be audible by the interlocutor to not disturb the others. When another

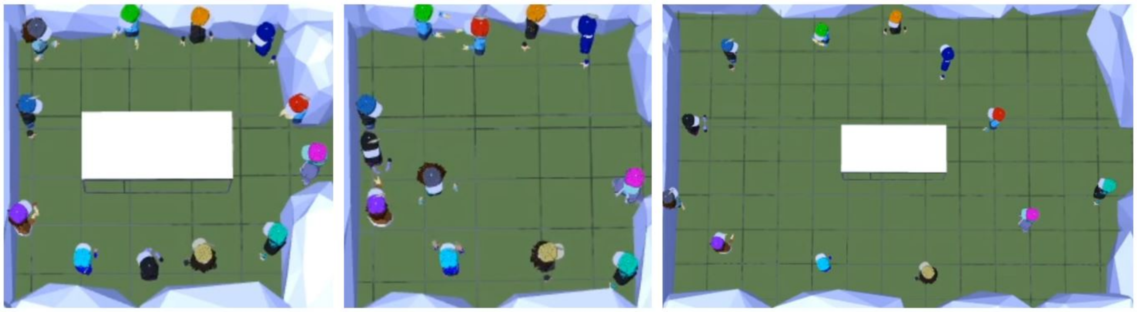

Fig. 2. Our observations in AltspaceVR: discussion situation with 12 people (left), several dialogues with two people each and one group of three people (middle), discussion situation with 12 people in a bigger room (right). Every participant is wearing a cap, which helps to identify the viewing direction.

person joins them, a discussion situation (3) arises. Here, we want to investigate for which distance and orientation the third person has to be involved in the conversation. Then, we want to introduce sound zones where only the voices of people one wants to follow can be heard (Figure 1, E-G) allowing multiple groups for people to communicate within one larger space. The affiliation to a certain zone can change over time. For instance, while listening to a lecture it may be necessary to communicate with the neighbor for an interactive part. Thus, different interactions are suitable for different proxemic zones [5].

3 APPROACH

It would be burdensome to change the sound zone every time manually in a menu. To improve VR communication, it is desirable that the VR system recognizes the communication situation and automatically changes the sound design. We intend to use machine learning to categorize the situations. For this, several important characteristics have to be analyzed: position, distance, and orientation [8].

4 FIRST EXPERIENCES

In our first experiments, we recognized that the more people were involved in a discussion, the greater the gaps. The distances were larger than we are used to in the real world, where we can perceive in a range of about 180° whether other people are present. Because the field of view (FoV) of Head-Mounted-Displays (HMDs) depends on the device specs and personal factors as eye positions, we measured the FoV which showed that all measured FoVs were less than 90°. In interviews, the participants reported that they had the desire to see all the group members in the FoV. Because the FoV in HMDs is smaller than in the real world, people took a step back so that they could see the others. As a result, they left the FoV of other participants, who then also stepped back. This process continued until everyone was near the wall (Figure 2 left and right). Such aspects which depend on the used hardware has to be taken into account when analyzing the conversation situation and to determine the necessary sound zone.

5 CONCLUSION

We conclude that Social VR needs sound concepts that automatically adapt to the respective conversation situation in the future. Our first results show that presented approach can improve the user experience in Social VR essentially. In the next steps, we want to take our concept further by implementing and evaluation it with users.

REFERENCES

[1] Michael Argyle and Janet Dean. 1965. Eye-contact, distance and affiliation. *Sociometry* (1965), 289–304.

[2] Jeremy N. Bailenson, Jim Blascovich, Andrew C. Beall, and Jack M. Loomis. 2003. Interpersonal Distance in Immersive Virtual Environments. *Personality and Social Psychology Bulletin* 29, 7 (2003), 819–833. https://doi.org/10.1177/0146167203029007002 PMID: 15018671.

[3] Paulo Bala, Raul Masu, Valentina Nisi, and Nuno Nunes. 2019. "When the Elephant Trumps": A Comparative Study on Spatial Audio for Orientation in 360º Videos. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3290605.3300925

[4] Mar Gonzalez-Franco, Antonella Maselli, Dinei Florencio, Nikolai Smolyanskiy, and Zhengyou Zhang. 2017. Concurrent talking in immersive virtual reality: on the dominance of visual speech cues. *Scientific Reports* 7, 1 (19 Jun 2017), 3817. https://doi.org/10.1038/s41598-017-04201-x

[5] Edward T. Hall. 1966. The hidden dimension. *Anchor Books* 14 (1966), 103–124.

[6] Adam Kendon. 2010. Spacing and Orientation in Co-present Interaction. In *Development of Multimodal Interfaces: Active Listening and Synchrony*, Esposito A., Campbell N., Vogel C., Hussain A., and Nijholt A. (Eds.). Springer, Berlin, Heidelberg, 1–15. https://doi.org/10.1007/978-3-642-12397-9_1

[7] Francisco Kiss, Sven Mayer, and Valentin Schwind. 2020. Audio VR: Did Video Kill the Radio Star? *Interactions* 27, 3 (April 2020), 46–51. https://doi.org/10.1145/3386385

[8] 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* (Cambridge, Massachusetts, USA) *(UIST '12)*. Association for Computing Machinery, New York, NY, USA, 13–22. https://doi.org/10.1145/2380116.2380121

[9] 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* (Hangzhou, China) *(CSCW '11)*. Association for Computing Machinery, New York, NY, USA, 445–454. https://doi.org/10.1145/1958824.1958893

[10] Miles L Patterson. 1977. Interpersonal distance, affect, and equilibrium theory. *The Journal of Social Psychology* 101, 2 (1977), 205–214.

[11] Martin S Remland, Tricia S Jones, and Heidi Brinkman. 1995. Interpersonal distance, body orientation, and touch: Effects of culture, gender, and age. *The Journal of social psychology* 135, 3 (1995), 281–297. https://doi.org/10.1080/00224545.1995.9713958

[12] Francesco Setti, Chris Russell, Chiara Bassetti, and Marco Cristani. 2014. F-formation Detection: Individuating Free-standing Conversational Groups in Images. *PLOS ONE* 10, 5 (sep 2014), e0123783. https://doi.org/10.1371/journal.pone.0123783 arXiv:1409.2702

[13] Aaron Willette, Nachiketa Gargi, Eugene Kim, Julia Xu, Tanya Lai, and Anıl Çamcı. 2020. Crossplatform and cross-reality design of immersive sonic environments. In *Proc. Int. Conf. New Interfaces Musical Expression (NIME '20)*. 127–130.

[14] Stephen Worchel. 1986. The influence of contextual variables on interpersonal spacing. *Journal of Nonverbal Behavior* 10, 4 (1986), 230–254. https://doi.org/10.1007/BF00987482