Accessibility of Co-Located Meetings

Accessibility of Co-Located Meetings

Introduction to the Special Thematic Session

Andreas Kunz[1(✉)], Reinhard Koutny[2], and Klaus Miesenberger[2]

[1] Innovation Center Virtual Reality, ETH Zurich, Zurich, Switzerland
kunz@iwf.mavt.ethz.ch
[2] Institut Integriert Studieren, Johannes Kepler University, Linz, Austria
{Reinhard.Koutny,Klaus.Miesenberger}@jku.at
https://www.icvr.ethz.ch,
https://www.jku.at/institut-integriert-studieren/

Abstract. Non-verbal communication is an important carrier of information. Even though the spoken word can be heard by blind and visually impaired persons, up to 60% of the overall information still remains inaccessibly to them due to its visual character [11]. However, there is a wide spectrum of non-verbal communication elements, and not all of them are of the same importance. In particular for group meetings, facial expressions and pointing gestures are relevant, which need to be captured, interpreted and output to the blind and visually impaired person.

This session first gives a systematic approach to gather the accessibility requirements for blind and visually impaired persons, from which two typical requirements are selected and discussed in more detail. Here, solutions for capturing and interpreting are provided, and finally the session introduces a concept for accessible user interfaces.

Keywords: Emotion recognition · Pointing gesture detection · Non-verbal communication

1 Introduction

Team meetings do not employ the spoken word only, but also heavily rely on visual information that is spatially distributed in a room, and to which sighted persons refer to by gesturing. Moreover, body language plays an important role for information exchange, but also for unconsciously managing a conversation. In particular, body language often refers to spatially distributed information such as different whiteboards or screens, but also to locations of other users (see Fig. 1). Thus, blind and visually impaired persons face two problems at the same time: (i) they lack important information for social communication, and (ii) they have to localize information that is spatially distributed.

A promising approach to overcome these issues could be a completely virtual meeting in a virtual environment, in which all participant are represented as avatars. Such a virtual meeting room would keep the spacial information distribution as this is helpful to the sighted users, but it also allows a blind and

K. Miesenberger et al. (Eds.): ICCHP-AAATE 2022, LNCS 13341, pp. 289–294, 2022.
https://doi.org/10.1007/978-3-031-08648-9_33

Fig. 1. Typical meeting room with multiple interaction spaces.

visually impaired person to retrieve information in the appropriate way e.g., by localisation movements using his or her smartphone, which would not be possible in a collocated meeting within a physical space.

However, such a virtual meeting room can only be as good as sensors allow capturing sighted persons' behavior such as facial expressions or referring gestures. Also, the further post-processing of the acquired data is important to avoid an overflow of information for the blind and visually impaired person on the one hand, but also to guarantee that no important information gets lost or modified on the other hand.

This special thematic session thus introduces a first overall approach if a digital accessible meeting room.

2 Accessibility Needs to Non-verbal Communication [15]

Enforced by the pandemics, many team meetings were held virtually, either through videoconferencing systems, or completely in a virtual environment such as the Metaverse using avatars [14]. It is most likely that such an IT-based communication will not disappear in the post-pandemic era. The supporting IT experienced a significant technological boost and is now able to support multi-user collaboration over the network in high quality. However, non-verbal communication and spatial information that is inherent to such team meetings might still be difficult to access by blind and visually impaired persons due to several reasons, e.g., low bandwidth that reduces audio fidelity, or too complex technical capabilities to capture and interpret non-verbal communication elements from the sighted persons. Consequently, a social interaction between sighted and visually impaired persons is hardly possible [5].

For such a social interaction - regardless whether it would be in a videoconferencing system or in a virtual environment - additional non-verbal cues

such as facial expressions or pointing gestures need to captured, interpreted and displayed to the blind and visually impaired person [12,13].

To support the most important non-verbal communication elements for social interaction in the future, the needs were acquired in a semi-structured interview with blind and visually impaired persons. Here, the most important elements were stated to be: (i) gaze and gaze direction, (ii) facial expressions, (iii) gestures, (iv) audio, and (v) touch. These communication channels - if not directly accessible - need to be captured, interpreted and displayed to the blind and visually impaired persons.

3 Facial Expressions and Emotion Recognition [9]

Facial expressions together with verbal intonation help to better understand a reaction or attention of a user in a meeting. They are thus important for a social interaction [2]. Regardless whether a team meeting is done through a videoconferencing system or in a virtual environment, facial expressions have to be recognized, categorized, and then displayed to the blind and visually impaired person. While emotions could also be detected from the voice pitch, i.e., from a frequency shift of the upper formant frequencies [16], this might not be possible because of a low bandwidth that does not allow transferring high frequencies, or simply because a user is doing facial expressions without speaking. Instead, a video signal from a user's webcam is taken and fed into a neural network such as a convolutional neural network [10]. Such networks are then trained with publicly available datasets such as AFEW[1] and deliver seven possible classes of facial expressions. However, these emotion classes are much too detailed for a team meeting and would overwhelm a blind and visually impaired person with information. In order to reduce the amount of emotion information, the seven different classes are clustered into the three categories "positive", "neutral", and "negative". After training the network with these new classes, facial expression taken with a regular webcam could be analyzed, and then the emotions were recognized with 97% (positive), 99% (neutral), and 64% (negative). The results could then either be displayed directly to the blind and visually impaired person, or they could be used to animate the facial expressions of an avatar in a virtual environment (see Sect. 2). To further reduce the information flow to the blind and visually impaired person, the "neutral" state could be neglected, since it doesn't give any further information to the spoken word.

4 Gesture Detection [3]

Within a conversation, sighted persons frequently use gestures to refer to objects in the nearby environment. Since these gestures cannot be accessed by blind and visually impaired persons, there is a need for detection and interpretation [4]. While also for gesture detection and interpretation deep learning approaches

[1] https://cs.anu.edu.au/few/AFEW.html.

exist, they are not applicable to referring gestures, e.g., on artifacts on a whiteboard, since here also the environment has to be taken into account. In particular for pointing gestures on a whiteboard's content, a high accuracy in detecting them is required to precisely display the corresponding artifact's content to the blind and visually impaired person [1,8]. However, the precision relies on various factors such as pointing accuracy, tracking accuracy, amount and position of artifacts on the whiteboard, etc. When referring to whiteboard content, mainly the three gestures "pointing", "pairing", and "grouping" are used. Mapping these gestures as trajectory on the whiteboard, the gestures mainly differ in their radii, in the angle between two succeeding tangents, and in the fact whether they enclose an artifact or not. Now, the user will interact for example in a virtual environment (see Sect. 2) on a virtual whiteboard (see Fig. 2).

Fig. 2. User performing a pointing gesture in a virtual environment.

Within the virtual environment shown in Fig. 2, in total 1350 different actions were performed, from which 44% were recognized correctly. The main reason for not detecting more gestures correctly is in the fact that users typically did not perform pointing gestures very accurately. However, even false recognition (26%) does not necessarily mean that it completely irritates the blind and visually impaired person, but it might just be a misinterpretation of a grouping by a pairing gesture or vice-versa, which still preserves the overall context. During the measurements, it never happened that the recognized gestures refer to a completely different region of the whiteboard.

5 User Interface Concept [6]

So far, the achieved information can be made accessible using standard interfaces such as a Braille display. However, in a group conversation there are also a lot of referring gestures to general spatial information in the nearby environment, which have to be made accessible to the blind and visually impaired persons. In some previous approaches, this spatial information was mapped on the

audio channel, which was then overloaded very quickly [7]. Thus, new approaches are proposed that offer a sufficiently high spatial resolution in order to localize objects in the near environment.

A first approach is to map the virtual meeting room on a web interface, on which blind and visually impaired persons could access important information by simply using a screen reader. In the web interface, e.g., the content of a digital whiteboard could be represented, as well as other users, their emotions and where they are pointing at.

A second approach employs a smartphone which translates the distance to an object, e.g., an artifact on a whiteboard, to vibration bursts with increasing repetition frequency the closer the smartphone comes to an artifact on the virtual whiteboard in front of the blind and visually impaired user. When selecting an artifact such as a card, the smartphone can be used to read out the content of this note, but also to manipulate the note e.g., by moving it to a different position on the virtual whiteboard.

A third approach also uses a vibration feedback to locate objects in the environment, but now the user wears a smartwatch as an output device.

The three proposed hardware settings are currently undergoing user studies for optimizing the overall system and to include feedback from blind and visually impaired users.

6 Summary

This special thematic session addresses a virtual collaboration environment that captures sighted users' facial expression as an indicator of their emotional state, and their pointing gestures. After an interpretation of the acquired measurements, the information is displayed on novel interface concepts to the blind and visually impaired person.

References

1. Dhingra, N., Valli, E., Kunz, A.: Recognition and localisation of pointing gestures using a RGB-D camera. In: Stephanidis, C., Antona, M. (eds.) HCII 2020. CCIS, vol. 1224, pp. 205–212. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-50726-8_27
2. El-Gayyar, M., ElYamany, H.F., Gaber, T., Hassanien, A.E.: Social network framework for deaf and blind people based on cloud computing. In: 2013 Federated Conference on Computer Science and Information Systems, pp. 1313–1319. IEEE (2013)
3. Gorobets, V., Merkle, C., Kunz, A.: Pointing, pairing and grouping gesture recognition in virtual reality. In: Computers Helping People with Special Needs, 18th International Conference; Joint Conference ICCHP-AAATE, Lecco, Italy, Proceedings. Springer (2022)
4. Kane, S.K., Wobbrock, J.O., Ladner, R.E.: Usable gestures for blind people: understanding preference and performance. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 413–422 (2011)

5. Kim, J.: VIVR: presence of immersive interaction for visual impairment virtual reality. IEEE Access **8**, 196151–196159 (2020)
6. Koutny, R., Miesenberger, K.: Accessible user interface concept for business meeting tool support including spatial and non-verbal information for blind and visually impaired people. In: Computers Helping People with Special Needs, 18th International Conference; Joint Conference ICCHP-AAATE, Lecco, Italy, Proceedings. Springer (2022)
7. Kunz, A., et al.: Accessibility of brainstorming sessions for blind people. In: Miesenberger, K., Fels, D., Archambault, D., Penáz, P., Zagler, W. (eds.) ICCHP 2014. LNCS, vol. 8547, pp. 237–244. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-08596-8_38
8. Liechti, S., Dhingra, N., Kunz, A.: Detection and localisation of pointing, pairing and grouping gestures for brainstorming meeting applications. In: Stephanidis, C., Antona, M., Ntoa, S. (eds.) HCII 2021. CCIS, vol. 1420, pp. 22–29. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-78642-7_4
9. Lutfallah, M., Käch, B., Hirt, C., Kunz, A.: Emotion recognition - a tool to improve meeting experience for visually impaired. In: Computers Helping People with Special Needs, 18th International Conference; Joint Conference ICCHP-AAATE, Lecco, Italy, Proceedings. Springer (2022)
10. Marinoiu, E., Zanfir, M., Olaru, V., Sminchisescu, C.: 3D human sensing, action and emotion recognition in robot assisted therapy of children with autism. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2158–2167 (2018)
11. Mehrabian, A., Ferris, S.: Inference of attitudes from nonverbal communication in two channels. J. Consult. Clin. Psychol. **3**, 248–252 (1967)
12. Oh Kruzic, C., Kruzic, D., Herrera, F., Bailenson, J.: Facial expressions contribute more than body movements to conversational outcomes in avatar-mediated virtual environments. Sci. Rep. **10**(1), 1–23 (2020)
13. Roth, D., Klelnbeck, C., Feigl, T., Mutschler, C., Latoschik, M.E.: Beyond replication: augmenting social behaviors in multi-user virtual realities. In: 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 215–222. IEEE (2018)
14. Tu, J.: Meetings in the Metaverse: Exploring Online Meeting Spaces through Meaningful Interactions in Gather. Town. Master's thesis, University of Waterloo (2022)
15. Wieland, M., Thevin, L., Machulla, T.: Non-verbal communication and joint attention between people with and without visual impairments. guidelines for inclusive conversations in virtual realities. In: Computers Helping People with Special Needs, 18th International Conference; Joint Conference ICCHP-AAATE, Lecco, Italy, Proceedings. Springer (2022)
16. Yildirim, S., et al.: An acoustic study of emotions expressed in speech. In: Proceedings of the Eighth International Conference on Spoken Language Processing (2004)