Non-verbal Communication and Joint Attention Between People with and Without Visual Impairments: Deriving Guidelines for Inclusive Conversations in Virtual Realities

Markus Wieland[1(✉)], Lauren Thevin[2], Albrecht Schmidt[1], and Tonja Machulla[1,3]

[1] LMU Munich, 80337 Munich, Germany
{markus.wieland,albrecht.schmidt}@ifi.lmu.de
[2] Université Catholique de l'Ouest, Angers, France
lthevin@uco.fr
[3] TU Dortmund University, 44227 Dortmund, Germany
tonja.machulla@tu-dortmund.de

Abstract. With the emergence of mainstream virtual reality (VR) platforms for social interactions, non-verbal communicative cues are increasingly being transmitted into the virtual environment. Since VR is primarily a visual medium, accessible VR solutions are required for people with visual impairments (PVI). However, existing propositions do not take into account social interactions, and therefore PVI are excluded from this type of experience. To address this issue, we conducted semi-structured interviews with eleven participants, seven of whom were PVI and four of whom were partners or close friends without visual impairments, to explore how non-verbal cues and joint attention are used and perceived in everyday social situations and conversations. Our goal was to provide guidelines for inclusive conversations in virtual environments for PVI. Our findings suggest that gaze, head direction, head movements, and facial expressions are important for both groups in conversations but often difficult to identify visually for PVI. From our findings, we provide concrete suggestions for the design of social VR spaces, inclusive to PVI.

1 Introduction

Social interactions are an emergent topic in virtual reality (VR). In 3D virtual spaces like Altspace VR, Rec Room and Horizon Worlds, people can interact with each other and also play games together. For realistic social interactions in VR the design of embodied virtual avatars is increasingly important [3]. This also includes non-verbal communication such as gaze, facial expressions, gestures, haptics, body posture, distance, and general noises. Today's technology makes it possible to integrate such non-verbal cues into 3D virtual spaces for better social interaction [2]. However, current VR applications have a strong focus

© Springer Nature Switzerland AG 2022
K. Miesenberger et al. (Eds.): ICCHP-AAATE 2022, LNCS 13341, pp. 295–304, 2022.
https://doi.org/10.1007/978-3-031-08648-9_34

on increasing the visual fidelity of such cues. As a result, people with visual impairments (PVI) profit only to a limited extent from these advances.

Various adaptations and tools have been proposed to increase the accessibility of VR for PVI. However, they focus on facilitating navigation through VR environments and the interaction with inanimate elements of the virtual scene but not on social interactions [4,8,15,17,18]. To successfully initiate social VR experiences, the user needs to be able to detect the presence of others and identify them if already acquainted. In real life situations, PVI employ a broad range of multisensory cues to achieve this task, some of which could already be integrated into present VR applications (for guidelines for inclusive avatars see [13]). Additionally, non-verbal communication cues, including those that facilitate joint attention, are an integral part of social interactions. Joint attention refers to the notion of drawing another person's attention to an object or situation. In persons without visual impairment, this is mostly achieved through gaze [9]. The communication through gaze has an evolutionary origin and initiating actions on the basis of others' gaze directions is more pronounced in humans than in other primates [5,14]. In real-world interactions, non-verbal communication and joint attention are likely to be used between persons with and without visual impairment. This should be reflected in virtual social experiences.

In the present work, we focus on non-verbal communication between PVI and others and how attention is jointly directed to objects or situations in real-world settings. A total of eleven people were interviewed in a semi-structured interview. Seven of them were PVI and four were their partners or friends who did not have a visual impairment. From the results, we propose guidelines for the design of representations of non-verbal communication and interaction cues inclusive to persons with and without visual impairment.

2 Related Work

Self-presentation is an important factor for social interactions in VR. Personalized avatars as well as realistic photogrammetry 3D scans are perceived as more human-like compared to an abstract avatar, resulting in a higher virtual body ownership [7,16]. Further, non-verbal cues in VR contribute to more engaging social interactions. Seeing facial expressions and bodily gestures of an avatar in VR lead to a more positive interaction experience [10]. Augmenting eye contact, joint attention, and grouping in social interactions in VR increased the perception of social presence [12]. In sum, combining realistic avatars with non-verbal communication signals can enhance the experience of an social interactions in VR. However, PVI do not profit from these types of non-verbal cues because they are designed specifically for visual perception. So far, attempts to increase VR accessibility for PVI focus on improving the interaction with the physical VR environment. To aid navigation, a cane was developed that provides a physical resistance when touching objects in VR, gives tactile and auditory feedback resembling real world sounds of the cane [17]. In another VR application for PVI the environment is generated entirely through sounds and PVI can walk through and perceive the environment [15]. SeeingVR provides tools such as a magnification lens, edge enhancement or text to speech that PVI can apply in VR [18]. PVI with macular degeneration can use a tool that increases the color and brightness in their central vision in VR [8]. The proposed studies help PVI navigate and explore virtual environments, but do not assist them in social interactions in VR.

3 Methods and Participants

We studied how social cues are used by PVI and partners without visual impairments in real-world scenarios, with the goal to propose implementations in VR. The study was conducted as a semi-structured interview, consisting of 22 questions for the PVI and 14 questions for their partners and friends. The questions were divided into the following six categories: gaze and head direction, facial expressions, gestures, sense of hearing, sense of touch, and joint attention. The respective categories asked whether this type of communication channel is used and how it is perceived. Seven PVI (four female, three male; average age of 56 years with a standard deviation of 23 years) and four sighted persons (one female, three male; average age 60 years with a standard deviation of 24 years) participated in the interview. The detailed overview about the PVI can be seen in Table 1. Note that except for P7, all participants are "blind" according to WHO classification [1]. Interviews were conducted online via Zoom in German and recorded with permission for further analysis. The interviews with the partners were conducted separately. All participants were financially compensated for their participation.

Table 1. Vision conditions of the seven PVI

ID	Sex	Age	Diagnosis	Residual Vision
P1	w	67	Inherited retinal degeneration	5%
P2	w	84	Age-related macular degeneration	5%
P3	m	48	Retinal detachment since birth	2%
P4	m	69	Myopia	2%
P5	w	29	Nystagmus since birth	3–5%
P6	m	69	Albinismus since birth	5%
P7	w	24	Stargardt	15%

4 Results

We grouped our results into six categories of non-verbal communication: gaze and head direction, facial expressions, gestures, sense of hearing, sense of touch, and joint attention. In each subsection, we first summarize the statements of the seven PVI, followed by supplementary statements of the four partners, and subsequently the guidelines for inclusive conversations for PVI in VR. Numerals in brackets indicate the number of participants who mentioned a specific theme.

4.1 Using and Perceiving Gaze and Head Direction

Participants with VI: Gaze, head direction, and head gestures are communicative cues that are difficult to perceive. The direction of the head is more perceptible than the gaze (5) than vice versa (1) and language is used to orient oneself in a conversation (1). Participants report that they consider gaze to be important (5) (P1, "It is theoretically important for me, but only very limited perceivable") and that they try to look their conversational partner in the eye (7). For example, P4 said "I do this for the reason that I signal readiness to receive". This is in spite of the fact that most cannot perceive their counterpart's eyes or gaze direction (6). All report to be familiar with situations where persons communicate through gaze alone but that they would typically use other cues such as giggling or the sound of clearing one's throat (4), movements (1) or another uncommon reaction (1) of their partner. One person stated not using alternative ways of communication. With regard to head direction, few persons are able to make use of this cue directly (2) and some are able to infer it from changes in the acoustics in their partner's speech, i.e., sound direction (4). Head gestures, such as nodding and shaking for negation, are perceivable by one person; all others reported that it depends on distance, context, and light conditions. It helped if these gestures were done more consciously and conspicuously than usual and accompanied verbally. One PVI also said that their close friends do not use head gestures during conversation, illustrating an adaptation of the social environment. All persons reported that the perception of gaze and head direction is influenced by light conditions (7) in a conversation. Lighting is in general problematic because too much light such as daylight or candles can be blinding (P6, "When there is a candle on the table, it blinds me.") and too little light worsens perception and the interlocutor has to talk more (5). Further, perception is poor when the interlocutor sits with their back to the window (2).

Sighted Partners: Partners stated that both parties always tried to look at each other during a conversation (4) and eye contact is important for them (3). One person said "I miss that, of course [eye contact with the partner]. When my partner is sitting across from me at the table, my partner doesn't see if I'm making faces if I'm laughing, angry, or crying anyway. My partner really does not see such things.", which shows the feeling of missing eye contact with the partner. One sighted partner said eye contact was important, while the PVI responded by not knowing for sure because it was not perceivable. Several sighted partners do not know whether the partner relies more on sight or on the head direction (3) whereas one sighted responded with head direction. In the perception of gaze and head direction under different lighting conditions, there are no differences in the PVI perceived by the sighted partners (4).

Guidelines for Inclusive Conversations: Eye contact is important for both groups in a conversation. Therefore, conversations in VR should provide multisensory cues based on gaze, head direction, and head movement. For example, when another person approaches a PVI for a conversation, an auditory cue could help the PVI to know that a person wants to talk to them and is trying to make eye contact. During a conversation, movements such as a head nodding could be represented with significantly larger animations and accompanied acoustically. Further, the virtual environment could offer adjustable light settings for single elements to adapt contrasts.

4.2 Using and Perceiving Facial Expressions

Participants with VI: PVI use facial expressions in standard ways and without conscious intention (7). In others, facial expressions are not perceivable in conversations (4) unless there is good lighting (1), the interlocutor is close (1) (P5, "When my partner does something funny with his face, he comes right in front of my face"), or well known (1). Emotions that are typically conveyed by the face are directly inquired for (4) or perceived through other cues such as voice, breathing, sighing, and posture (3) (P5, "For example, if someone becomes very emotional, the person speaks more slowly or pauses more when speaking, I notice that. This means that the person may not be doing so well at the moment.").

Sighted Partners: The use of facial expressions during a conversation is confirmed by the sighted partners (4). They themselves use their facial expressions in conversation with their partner as with all other conversational partners (2), to a lesser degree (1), or not at all (1).

Guidelines for Inclusive Conversations: Facial expressions are mostly not perceivable to PVI, so they use other cues. Thus, the information should be communicated with multisensory cues. For example, the facial expression of the conversation partner could be recognized and then played back with audio cues or joyful facial expressions are accompanied by soft background music. Further, an automatic magnifier could be integrated that enlarges the face of the conversation partner and displays it in good lighting conditions.

4.3 Using and Perceiving Gestures

Participants with VI: The use of gestures (6) can be divided into semaphoric, deictic, and gesticulation gestures. Some PVI use various forms of gestures and therefore multiple mentioning was possible. Semaphoric gestures are hand gestures that stand for specific meanings. They are occasionally used, such as in the form of thumbs up (2) and semaphoric emoticons on a smartphone (1) (P3, "Some gestures I can imagine a certain way based on this description [screen reader] alone."). Deictic gestures, i.e., pointing, are only used when the location is well known (2) (P5, "[...] if it is a place I know well, then I point to it. But I think I say that in the context that, it is on the cabinet."). Further, gesticulations are used to express themselves in a conversation (2) and one of the PVI learned the concept of gesticulation from an early teacher. One PVI does not use gestures at all. Several factors have an influence on the perception of gestures and multiple mentioning was possible: the bigger, expressive, and flailing the gesture, the better it can be perceived (4) (P1, “The bigger they are [gestures], the better I can perceive them.”), light conditions and the change of colors through the light during a gesture (2), distance (2), air flow (2), context (1), and moving fabric of clothes (1). The recognition of two different gestures in one movement (e.g., semaphoric and then deictic) depends on distance, light condition and whether the attention is focused on the interlocutor (5), if the object for a deictic gesture is known in advance (1), would be ignored unless the interlocutor expresses it verbally (1).

Sighted Partners: The use of gestures in conversation with the partner is divided into normal use (2), less use (1) and no use at all (1). Gesture recognition for the PVI partner is only possible under the following conditions: when the partner is nearby (2), when something is pointed at and accompanied verbally at the same time (1). One sighted person mentioned that the partner is not able to recognize gestures.

Guidelines for Inclusive Conversations: Recognizing gestures for PVI is possible under certain conditions. In order to aid PVI to recognize gestures, the virtual environment should enhance semaphoric or deictic gestures with a larger representation of the gestures and with accompanying sounds. The connection between a deictic gesture to an object could be indicated with a high-contrast ray and announced verbally. Enhanced gesticulation could be used to personalize the avatars of social partners for easier identification.

4.4 Using Sense of Hearing

Participants with VI: Hearing is used for different tasks in a conversation: recognize behavior (4), recognize mood (1) (P5, “It is also important how the voice changes. For example, I can hear how someone smiles, so I know how the voice changes if I know the person.”), to draw conclusions about the interlocutor (1), and to infer the posture of the interlocutor, since people speak differently in different postures, and then in turn infer behavior (1). Non-communicative sounds of the interlocutor in a conversation are perceived as distracting and disturbing (3), or as interesting, because conclusions about the interlocutor can be drawn from it (1). However, general noises are not paid attention to (2) and it also depends on the context of the conversation (1).

Sighted Partners: The participants reported that they do not use consciously sounds in a conversation with their partners (4) and they do not consciously change their voice to add more emotion to what they say (3), except for one participant.

Guidelines for Inclusive Conversations: Since the voice in a conversation can provide important information about the behavior of the interlocutor, it should be possible to eliminate disturbing noises or background sounds. However, for people who consciously use sounds of the conversation partner to find out about behavior, sounds such as footsteps or a nervous shaking should be integrated or presented as a multisensory cue.

4.5 Using Sense of Touch

Participants with VI: The sense of touch is used in conversations only with friends (3) and when socially appropriate (1), to find out how serious the communication is meant through hugs and handshakes (1) (P4, "what I just always want to know is how serious it is meant, and that is also important for further communication"), or to attract attention in a conversation (1). One person state not to use the sense of touch in social interactions. When participants were asked if they were touched in a conversation and how they felt about it, they reported that it only occurs with friends (3), and it is fine as a sign of attention (1). Whereas it is sometimes annoying when people approach the PVI and ask if help is needed (1) (P6, "I sometimes have problems keeping people off my back. They think I have bad eyesight, and then they always come close to me.") and it also occurs that no touch at all takes place during a conversation (2).

Sighted Partners: The sense of touch is used according to general social norms (2) or to a lesser degree (2) in conversation with the partner.

Guidelines for Inclusive Conversations: The sense of touch is used socially appropriately and is also expected from the interlocutor. In a virtual environment, functions can be used that allow touch only after the person being touched has given permission or the function could be activated for friends so that they can touch each other or approach each other closely.

4.6 Joint Attention

Participants with VI: When participants are handed an object, it is recognized verbally (3) and by the movement of the body or arm (3) (P1, "I would perceive it visually. I think the person can only give it to me if the person sits close enough to me.") or by placing the object directly in the hand (1), but it is also difficult to recognize without verbal cues (1). The situation in which the participants are made aware of objects or situations in their environment usually happens with verbal cues (3), but this does not occur often because they know where everything is located in their environment (2) (P5, "In my apartment, I just know where everything is.") and objects are often placed directly in front of them (1). In public, the participants are verbally made aware of objects or situations (7), and further they indicated that they are additionally made aware by touch (2). If participants wanted to draw the attention of others to an object or situation in their environment, it happens verbally (4), with gestures because the partner no longer hears well (1), with deictic gestures but only if the location of the object is known (1) and getting closer to the object (1) (P7, "I do not really point at objects that you can hold in your hand. Instead, I usually get very close to the object I want to show or draw attention to.").

Sighted Partners: The partner's attention to an object or situation is drawn via verbal cues (3) and this does not occur often in the home environment because the partner knows where the objects are located (1). In addition, it was mentioned that the partner's attention is unintentionally drawn to something by making noises (1) ("When I drink in the kitchen, my partner hears that too and asks for a glass of water or something."). The sighted partner are made aware of objects or situations by their partners through verbal cues (2) and if the partner knows where the object is with a deictic gesture, otherwise verbally (1). However, it was also reported that one's attention is not drawn to an object or situation by one's partner because sighted people can see everything (1) ("My partner cannot direct my attention to any objects. It only works the other way around."). In terms of a difference in non-verbal cues when talking to their PVI partner or a sighted person, it was mentioned that less gestures are used (2) ("I use fewer gestures because I know they are ineffective with my partner."), no eye contact although it is important for them as a couple (1) and communication is generally more verbal, but this is usually done unconsciously (1) ("Communication shifts to the verbal level, but this happens rather unconsciously. You just adapt and somehow it becomes automatic without thinking about it.").

Guidelines for Inclusive Conversations: Verbal cues to objects or situations are necessary for joint interactions. To focus joint attention on an object or situation, short automatic verbal cues could be given. In addition, a high-contrast visual and auditory signal related to the PVI head direction and the sighted interlocutor's gaze could signal that both are attending to the same object in a collaborative interaction to establish joint attention. Further, verbal cues to an object or situation could automatically lead to a deictic gesture of the PVI's avatar toward the object or situation.

5 Discussion

Here, we presented the results of an interview with eleven participants regarding non-verbal communication and joint attention. The findings are used to propose guidelines for inclusive conversations in virtual environments for PVI. Our recommendations and examples include translating behaviors, verbalizations, and appearances into different formats. VR offers the opportunity to make various visual cues perceivable to PVI in a conversation. For example, by integrating facial expressions[1] into social VR environments, they can be made perceivable

[1] www.vive.com/us/accessory/facial-tracker/.

through other sensory cues, establishing bidirectional communication. However, the wide range of possibilities offered by VR also raises privacy issues. Both interlocutors need to know that their behavior, verbalizations, and appearances translate into different formats and commit to these inclusive conversations. Future work should address how to make VR accessible to PVI by also transferring existing real-world aids, such as eye-gaze glasses that provide tactile feedback when viewing at a PVI, to VR [11] or augmenting facial expressions [6]. In addition, a set of tools could be provided for inclusive conversations in VR that can be automatically integrated into an existing VR application similar to the approach of the toolbox SeeingVR [18]. Our future research will address the development and evaluation of different supporting methods for non-verbal signals for PVI in VR.

Acknowledgments. Research was supported by the BMBF (project HIVE: grant no. 16SV8183). We would also like to thank the reviewers and ACs for their work and valuable feedback.

References

1. Blindness and vision impairment: World Health Organization (2021). https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment
2. Cha, H.S., Choi, S.J., Im, C.H.: Real-time recognition of facial expressions using facial electromyograms recorded around the eyes for social virtual reality applications. IEEE Access **8**, 62065–62075 (2020)
3. Freeman, G., Maloney, D.: Body, avatar, and me: the presentation and perception of self in social virtual reality. Proc. ACM Hum.-Comput. Interact. **4**(CSCW3), 1–27 (2021)
4. Kim, J.: VIVR: presence of immersive interaction for visual impairment virtual reality. IEEE Access **8**, 196151–196159 (2020)
5. Kobayashi, H., Kohshima, S.: Unique morphology of the human eye and its adaptive meaning: comparative studies on external morphology of the primate eye. J. Hum. Evol. **40**(5), 419–435 (2001)
6. Lang, F., Schmidt, A., Machulla, T.: Augmented reality for people with low vision: symbolic and alphanumeric representation of information. In: Miesenberger, K., Manduchi, R., Covarrubias Rodriguez, M., Peňáz, P. (eds.) ICCHP 2020. LNCS, vol. 12376, pp. 146–156. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58796-3_19
7. Latoschik, M.E., Roth, D., Gall, D., Achenbach, J., Waltemate, T., Botsch, M.: The effect of avatar realism in immersive social virtual realities. In: Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology, pp. 1–10. ACM, Gothenburg Sweden (2017)
8. Masnadi, S., Williamson, B., Gonzalez, A.N.V., LaViola, J.J.: VRiAssist: an eye-tracked virtual reality low vision assistance tool. In: 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), pp. 808–809. IEEE, Atlanta, GA, USA (2020)
9. Mundy, P., Newell, L.: Attention, joint attention, and social cognition. Curr. Dir. Psychol. Sci. **16**(5), 269–274 (2007)

10. 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), 20626 (2020)
11. Qiu, S., Hu, J., Han, T., Osawa, H., Rauterberg, M.: An evaluation of a wearable assistive device for augmenting social interactions. IEEE Access **8**, 164661–164677 (2020)
12. 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, Tuebingen/Reutlingen, Germany (2018)
13. Thevin, L., Machulla, T.: Guidelines for inclusive avatars and agents: how persons with visual impairments detect and recognize others and their activities. In: Miesenberger, K., Manduchi, R., Covarrubias Rodriguez, M., Peňáz, P. (eds.) ICCHP 2020. LNCS, vol. 12376, pp. 164–175. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58796-3_21
14. Tomasello, M., Hare, B., Lehmann, H., Call, J.: Reliance on head versus eyes in the gaze following of great apes and human infants: the cooperative eye hypothesis. J. Hum. Evol. **52**(3), 314–320 (2007)
15. Torres-Gil, M.A., Casanova-Gonzalez, O., Gonzalez-Mora, J.L.: Applications of virtual reality for visually impaired people. WSEAS Trans. Comput. **9**(2), 184–193 (2010)
16. Waltemate, T., Gall, D., Roth, D., Botsch, M., Latoschik, M.E.: The impact of avatar personalization and immersion on virtual body ownership, presence, and emotional response. IEEE Trans. Vis. Comput. Graph. **24**(4), 1643–1652 (2018)
17. Zhao, Y., et al.: Enabling people with visual impairments to navigate virtual reality with a haptic and auditory cane simulation. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp. 1–14. ACM, Montreal, QC, Canada (2018)
18. Zhao, Y., Cutrell, E., Holz, C., Morris, M.R., Ofek, E., Wilson, A.D.: SeeingVR: a set of tools to make virtual reality more accessible to people with low vision. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–14. ACM, Glasgow, Scotland, UK (2019)