

Accessible Nonverbal Cues to Support Conversations in VR for Blind and Low Vision People

Crescentia Jung*
cj382@cornell.edu
Cornell University
New York, NY, USA

Jazmin Collins*
jc2884@cornell.edu
Cornell University
New York, NY, USA

Ricardo E. Gonzalez Penuela
reg258@cornell.edu
Cornell University
New York, NY, USA

Jonathan Isaac Segal
jis62@cornell.edu
Cornell University
Ithaca, NY, USA

Andrea Stevenson Won
asw248@cornell.edu
Cornell University
Ithaca, NY, USA

Shiri Azenkot
shiri.azenkot@cornell.edu
Cornell Tech
New York, NY, USA

ABSTRACT

Social VR has increased in popularity due to its affordances for rich, embodied, and nonverbal communication. However, nonverbal communication remains inaccessible for blind and low vision people in social VR. We designed accessible cues with audio and haptics to represent three nonverbal behaviors: eye contact, head shaking, and head nodding. We evaluated these cues in real-time conversation tasks where 16 blind and low vision participants conversed with two other users in VR. We found that the cues were effective in supporting conversations in VR. Participants had statistically significantly higher scores for accuracy and confidence in detecting attention during conversations with the cues than without. We also found that participants had a range of preferences and uses for the cues, such as learning social norms. We present design implications for handling additional cues in the future, such as the challenges of incorporating AI. Through this work, we take a step towards making interpersonal embodied interactions in VR fully accessible for blind and low vision people.

CCS CONCEPTS

• **Human-centered computing** → **Accessibility**.

KEYWORDS

blind, low vision, VR, accessibility

ACM Reference Format:

Crescentia Jung, Jazmin Collins, Ricardo E. Gonzalez Penuela, Jonathan Isaac Segal, Andrea Stevenson Won, and Shiri Azenkot. 2024. Accessible Nonverbal Cues to Support Conversations in VR for Blind and Low Vision People. In *The 26th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '24)*, October 27–30, 2024, St. John's, NL, Canada. ACM, New York, NY, USA, 13 pages. <https://doi.org/10.1145/3663548.3675663>

1 INTRODUCTION

Social virtual reality (VR) is growing in popularity but remains inaccessible to blind and low vision (BLV) people. Social VR has thousands of daily users that participate in a variety of social activities from informal parties to remote business meetings. In addition,

the embodied nature of social VR allows people to communicate with nonverbal behaviors such as gestures and facial expressions. However, these behaviors are generally rendered visually, and this limits BLV people's access to social information, excluding them from full participation.

Researchers have explored how to make social VR more accessible for BLV people [11, 17, 23, 52, 54]. For example, Zhang et al. explored avatar diversity and the importance of self-presentation of people with disabilities in social VR [52]. Additionally, Collins and Jung et al. studied the use of a sighted guide as an accessibility tool in social VR [11].

However, there has been little exploration of making nonverbal behaviors accessible in social VR. Some have investigated what kinds of behaviors BLV people wish to be made accessible in VR, as well as initial guidelines for the best ways of making certain behaviors accessible. For example, Wieland et al. [50] conducted a user study to identify which nonverbal behaviors BLV people used in conversations with sighted partners and should be carried over into social VR, such as facial expressions, head movements, and gaze. Similarly, Collins and Jung et al. [10] worked with a BLV person to identify best practices for representing gaze in VR. Ji et al.'s [23] "VRBubble" used audio beacons to indicate proxemic information about avatars to improve BLV people's awareness of others.

Prior works such as those by Collins and Jung et al. [10] and Ji et al. [23] have designed accessibility features for one type of nonverbal behavior at a time. The participants evaluating these systems also experienced only limited, pre-recorded conversations with agent-avatars, instead of actual people. This is understandable, considering the complexity and the immense range of the nonverbal behaviors real people use in conversation. Trying to design representations of multiple behaviors at a time is challenging, and may involve various problems such as determining which behaviors to represent, the best ways to represent them, and ensuring all representations can be used at once without becoming overwhelming. However, only having one accessible behavior during pre-recorded interactions does not represent real conversations, or provide enough nonverbal information to fully support conversation. BLV users need access to multiple nonverbal behaviors to support conversations with others in VR.

To address this need, we sought to design multiple accessible nonverbal cues that could be used simultaneously in social VR for

*Crescentia Jung and Jazmin Collins are co-authors and contributed equally to this research.

BLV people. Specifically, we aim to address the following research question: **What set of accessible nonverbal cues can support BLV users in conversations with others in VR?** We use the word “cue” here and in the remainder of our paper to refer to a representation of a nonverbal behavior, while the word “behavior” refers to the nonverbal behavior itself.

We conducted a user-centered design process to design accessible audio and haptic cue representations of nonverbal behaviors. We iterated on five cue designs with six BLV participants. At the end of the process, we found that five cues were overwhelming for participants, and narrowed our focus to three final cues: eye contact, head nodding, and head shaking.

We then evaluated this set of three cues with 16 participants in social VR conversation tasks. Each participant completed two conversation tasks: one with cues and one without. We found that our nonverbal cues significantly improved participants’ accuracy in detecting certain social behaviors about their conversation partners without distracting them from the conversation. Participants also reported feeling more confident about “reading the room”, especially knowing how much attention they were receiving. Our discussions with participants yielded design implications for how nonverbal behaviors might be represented accessibly in the future, including more behaviors participants wanted to access, and potential ways of managing larger sets of accessible cues for these behaviors.

Our contributions include the design and evaluation of three accessible nonverbal cues in VR and novel design implications for sets of accessible nonverbal cues.

2 RELATED WORK

2.1 Accessible VR for BLV People

VR experiences tend to be over-reliant on visual information, making them inaccessible for BLV people. To address this, researchers have explored how to make virtual environments and the content within them perceivable to BLV users. Much of this work has focused on providing *environmental* information via haptics or audio to enhance BLV users’ individual experiences in VR, but far less has focused on providing *social* information to enhance social experiences with other people.

Many researchers have investigated how to provide information about a VR environment’s layout to support BLV people’s navigation [2, 15, 18, 19, 34, 43, 53, 54]. For instance, Andrade et al. [2] designed EchoHouse, a virtual environment where users navigate with “echolocation”, using real-time audio cues emitted from objects in the space. More recently, Collins and Jung et al. [11] proposed supporting navigation through sighted guides. In their work, they implemented a system that allowed a BLV user to pair up with a sighted human guide to do visual interpretation of and navigate any virtual space together.

Besides navigation, researchers have explored ways of conveying details about objects in virtual environments through audio and haptics [16, 25, 36, 42, 55]. Zhao et al. [55] created a developer toolkit called SeeingVR, which offers 14 different visibility enhancements to make VR experiences accessible for users with low vision (e.g., magnification, highlighting salient objects, font size adjustment, and others). Researchers have also developed approaches using haptics and gestures for blind users, such as Penuela et al.’s [16]

haptic gloves allow BLV users to perceive the shape of objects through force feedback and elicit descriptions about objects in the environment through gestures.

Some researchers have also created accessible VR games that use nonvisual modalities to help BLV users complete complex objectives in virtual environments [4, 13, 14, 30, 33, 35, 47]. For example, Wedoff et al. [47] designed Virtual Showdown, a VR audio game that provides spatialized audio cues to help players find, move towards, and hit a ball back to an opponent.

While the above efforts have taken an important step towards making *individual* VR experiences accessible for BLV users, they do not account for the social aspects of environments where multiple users are present, like social VR applications. These applications are often designed so users interact with each other to complete shared objectives or bond with each other in conversations. In such cases, communicating social information about other people is more important to support BLV users’ abilities to participate in the virtual space. Without access to social information about other people, BLV users may remain isolated in social VR spaces, even with accessibility enhancements for individual experiences. Our work seeks to address this gap by supporting BLV people’s access to a key aspect of communication in social VR: nonverbal behaviors.

2.2 Nonverbal Behaviors in Social VR

Nonverbal behaviors are an important aspect of social VR and have been explored extensively in prior work. However, this work has largely focused on understanding how they are used, rather than how to make them accessible.

Researchers have explored which nonverbal behaviors are most commonly used in social VR [31, 45]. Maloney et al. [31] observed people’s use of nonverbal behaviors on social VR applications and conducted an interview study of participants’ perceptions of nonverbal communication in social VR. The authors uncovered several types of nonverbal communication methods used in social VR, including applauding, dancing, and even flying or using emojis. Among the most common across platforms were hand gestures and head movements, like waving or nodding. Similarly, Tanenbaum et al. [45] explored ten popular social VR platforms and took inventory of their existing designs for expressing nonverbal behaviors. While the quality of each platform’s support of these behaviors varied, they found that most social VR platforms included the ability to express behaviors such as proxemics, facial expressions, and gestures.

Other researchers have explored particular types of nonverbal behaviors, including gestures, nodding, and eye gaze, and how they affect perceptions of conversation partners in VR [1, 21, 24, 26, 28, 51]. For instance, Aburumman et al. [1] conducted VR studies with 21 participants, having participants experience two types of nodding behavior in a conversation with agent-avatars. They found that an agent-avatar that nodded while participants spoke resulted in higher levels of trust felt towards the agent-avatar. Ide et al. [21] observed the effect of symbolic gestures to invoke gestures on virtual avatars (e.g., thumbs-up emoji to do a thumbs-up, surprised emoji to make a surprised face) on brainstorming tasks. They found that symbolic gestures helped participants express their emotions and supported social communication. Similarly, Kurzweg et al. [26]

explored the effect of body language (e.g., crossing arms, drinking) on communication in VR and found that body language can indicate willingness to communicate and attentiveness towards others in a virtual environment.

These investigations demonstrate the importance of nonverbal behaviors for communication in VR. Despite this, there has been little exploration of making these behaviors accessible to support BLV communication needs [10, 23, 41, 49, 50]. One of the most notable efforts is Wieland et al.'s [50] work on identifying important nonverbal behaviors in social VR. Wieland et al. conducted interview studies with eleven participants, seven BLV and four sighted companions of BLV people, to identify which nonverbal behaviors they used most often in conversation, and which should be carried over into VR for effective communication. They found that gaze, head direction, head movements, and facial expressions were all important to identify, though they could be difficult to notice for BLV participants.

Other researchers have developed design guidelines for making certain nonverbal behaviors accessible. Collins and Jung et al. [10] probed the design space of accessible gaze feedback in VR with a blind co-designer. The co-designer experienced and adjusted feedback for two kinds of gaze (mutual and resting gaze) through 5 parameters (e.g., Modality, Strength, Duration, etc.) to create their preferred gaze feedback in a VR prototype. Their study concluded with recommendations to send haptic vibrations to a blind user via handheld controllers when someone was looking at them in VR. Ji et al. [23] explored avatar proxemics in VR. They designed a system where audio beacons indicated proxemic information and tested it in user studies with 12 BLV participants. The system improved BLV people's awareness of others, providing concrete design recommendations for representing proxemics in social VR.

Both of these works have introduced initial approaches to making nonverbal behaviors accessible to BLV people and are an important step in supporting communication. However, they each only focus on one type of nonverbal behavior at a time, namely eye gaze and avatar proxemics. People use multiple nonverbal behaviors (e.g., eye gaze, head movements, and gestures) when engaging in a social space. As stated by Wieland et al. [49, 50], BLV people want access to multiple nonverbal behaviors that are used in conversation. In order to effectively support BLV people's conversation needs, multiple nonverbal behaviors should be made accessible at once. In addition, these works conducted their user evaluations with pre-recorded agent-avatars in scripted conversations, rather than supporting real conversations with unpredictable human conversation partners. These limitations represent a significant gap in current research efforts to support communication in VR for BLV people. Our work seeks to address this gap by implementing multiple nonverbal cues in conversations in VR to evaluate how these accessible representations of nonverbal behaviors support real conversations with others.

3 DESIGNING ACCESSIBLE NONVERBAL CUES

Our goal was to design a set of robust accessible nonverbal cues that could support conversations in VR. To achieve this, we conducted an iterative user-centered design process to convey a set of nonverbal

behaviors via accessible nonverbal cues. We worked in a mixed-ability team with professionals experienced in designing accessible technology. We also conducted a series of early informal design sessions with BLV co-designers. Throughout the design process, we met weekly to discuss and test prototypes.

To guide our design process, we developed four design considerations based on: (1) discussions among our mixed-ability team, (2) informal design sessions with BLV co-designers, and (3) past work in designing accessible representations of nonverbal cues in VR [10, 23].

- **Cues should leverage nonvisual modalities supported by VR technology.** We should consider the nonvisual capabilities of commercial VR technologies such as haptic vibrations, audio patterns, and spatialized audio.
- **Audio cues should represent the emotions behind each nonverbal behavior.** Leveraging the versatility of audio, audio cues should evoke emotions that match the behaviors they are representing, like positive emotions for a smile.
- **Cues should be unobtrusive during a conversation.** Cues should not be too loud, long, or distracting to the point where they interrupt or make conversation difficult for users.
- **Cues should be understandable when multiple cues are being used simultaneously.** Since people use multiple nonverbal behaviors simultaneously in conversation, the cues should be playable simultaneously, without overwhelming the user.

After finalizing these design considerations, we needed to establish which nonverbal behaviors to make accessible. We first considered the broad scope of possible nonverbal behaviors identified by prior work and narrowed them down to a set that is useful in conversations for BLV people. Wieland et al. [49, 50] found that eye gaze, head direction, head movements, and facial expressions were important for BLV people during conversations. When considering different forms of eye gaze, Collins and Jung et al. [10] and Wieland et al. [49, 50] found that eye contact information was the most preferred and important for BLV people. Maloney et al. [31] noted that some of the most commonly used nonverbal behaviors in VR included nodding and head shaking. Finally, existing research on facial expressions lists smiling and frowning as two of the most common expressions people make [20]. Considering these works, we narrowed our focus to five nonverbal behaviors to design as accessible cues: eye contact, head nodding, head shaking, smiling, and frowning.

3.1 Design Process

To create our initial designs, we examined existing sound libraries used to indicate information. After sorting through a variety of libraries (e.g., Facebook's emoji sound effects [12], vocal bursts such as laughs and sighs [39]) we created our initial designs of cues for the five nonverbal behaviors using Paquette et al.'s database of musical emotional bursts. Each burst in this dataset was a brief music clip that Paquette et al.'s participants associated with specific emotions [38].

We sought to iterate on the initial designs of our cues with BLV users to ensure they would be usable and understandable. To do

this, we conducted studies with six BLV participants. Each participant took part in a single in-person 90-minute session where they experienced the most recent iteration of our cue designs. We wanted to demonstrate our cue designs to participants in various scenarios, both in the physical world and in VR, to give them a good idea of what having conversations with these cues would be like. Thus, each study session contained four parts: (1) an introduction to the nonverbal cues in the physical world, (2) a one-on-one conversation augmented by the cues in the physical world, (3) a one-on-one conversation augmented by the cues in VR, and (4) a three-person group conversation in VR.

We represented the cues for participants in different ways in the non-VR and VR settings. In our non-VR introduction to the cues, we played each of the cues (eye gaze, head nodding, head shaking, smiling, or frowning) one by one from a laptop to share the cue audio while participants held VR controllers to receive haptic feedback. In the non-VR conversation, participants again held controllers to receive haptic feedback, but the researcher who was speaking to them played the audio cues from a phone in their hand, so that participants would hear the audio coming from the direction of the person they were speaking with. Finally, for the VR settings, we developed a VR prototype where researchers and the participant would enter a multi-user virtual environment together. Within this environment, the researchers could manually trigger nonverbal cues when they performed certain nonverbal behaviors to augment conversations. Participants used a VR headset and controllers to enter the scene and heard audio cues spatialized to the researchers' avatars while feeling haptic feedback from cues in their controllers.

We wanted to improve our cue designs with participants and allow them to try any suggestions they had for the cues during their study session. To do this, we asked participants for their perceptions of each cue after the conversations, including emotion invoked, level of distraction, and suggestions for improvements. After discussing the initial cue designs, we worked with participants to develop new iterations of the cues based on their critiques. Participants could request changes such as the length of the haptic cue, the volume of the audio, and the type of audio file being played (e.g., a musical note, a TV show sound effect, a recorded laugh, etc.). Participants could also create their own audio cues from scratch by directing the research team to create new sound effects. After making changes, we would demonstrate the new cues for participants, prompt them for feedback, and continue iterating. All of the changes we made to the cues were implemented via a laptop running Unity and the audio mixing software Garageband. Participants experienced the new cue iterations by listening to audio played from the laptop and holding VR controllers to receive haptic feedback. At the end of the session, we had participants comment on which iterations of the cues they preferred the most. We reached consensus for final designs when three participants in a row preferred the current versions of the cue designs without requesting modifications.

Following these sessions, we developed three additional design considerations from participants' feedback:

- **Accessible audio cues should be short for usability and accuracy.** Shorter lengths for the accessible nonverbal cues were preferred. If cues are too long, they become disruptive

to the conversation (e.g., a two-second clip representing head shaking drowned out conversation and tended to last over a second after head movements stopped).

- **Familiar sound effects are preferable for inciting emotions.** Sound effects that imitated sounds from media sources, like television game shows, were easier to associate with specific emotions than musical patterns.
- **Haptics should be used to represent frequent nonverbal cues.** Since audio repeating continuously would disrupt conversation, cues such as eye contact that occur frequently should be represented by less-disruptive haptic vibrations.

3.2 Final Designs

After receiving consistent feedback from participants that cues were easy to understand and unobtrusive to conversation, we stopped iterating on our designs. The final designs were as follows ¹:

- **Eye Contact.** A continuous haptic buzz that lasted for the duration of the eye contact.
- **Head Nod.** A succession of one high-pitched note followed by a slightly lower note from a xylophone, repeating twice and lasting 1 second.
- **Head Shake.** A succession of one low-pitched note followed by a slightly lower note from a flute, repeating twice and lasting under 1 second.
- **Smile.** A high-tone chime-like sound effect, formed of a series of cheerful ringing notes lasting 2 seconds.
- **Frown.** A low-tone trumpet sound effect lasting for 1 second.

An overarching takeaway we found with these cues was that **participants found a set of five cues was overwhelming to learn and immediately use in conversation**. Thus, we decided to **explore a smaller set** of nonverbal cues, selected from these five. We selected eye gaze for this subset since our participants responded the most positively to eye contact and its non-intrusive, haptic form. We also chose the two head movement cues—head nodding and head shaking—since they are more commonly used in mainstream VR than facial expressions due lack of integration with face-tracking technology.

3.3 Discussion and Implications

Our goal for the design process was to iteratively design accessible versions of nonverbal cues. We had not yet tested how effective these designs were at conveying social information about a conversation partner. One type of social information that many of our participants were particularly interested in was attention, specifically, in “reading the room” to know if their partner was paying attention to them. However, as we had focused primarily on determining how disruptive or intuitive the cue designs were, we had not examined the cues' abilities to convey information like this. Further work was required to specifically evaluate how well cues can convey a conversation partner's level of attention.

We had also only tested the cues with a small set of co-designers who were used to experimenting with and designing new accessible technologies. While their feedback was useful for designing effective nonverbal cues, it did not represent how easy these cues would

¹All tracks for audio cues are provided here: <https://soundcloud.com/shadowdios/sets/nvc-study-sound-effects>

be to learn and utilize for other BLV users. To truly understand how useful our cues would be, we needed to evaluate them with a larger and more diverse group of BLV participants.

4 EVALUATING THE CUES IN CONVERSATIONS

We aimed to evaluate the cues' effectiveness in real-time conversations in VR. To do this, we needed to consider what it meant for nonverbal behaviors to support conversations. Nonverbal behaviors convey a wide range of social information, from emotional reactions to level of agreement. For example, someone may look at you to indicate that they are listening or nod their head in agreement with what you are saying. Our previous design process had established our participants' interest in determining the attention levels of their partners. Since our set of three nonverbal cues, eye contact, head nodding, and head shaking, also commonly represent attention [8, 27, 32, 44], we selected attention as our focus for the evaluation. This would allow us to see if our cues could support our prior participants' desired conversational needs and whether our cues could convey important social information to support conversations.

We designed our evaluation with certain key questions in mind:

- How *accurate* are participants at detecting attention with and without accessible nonverbal cues?
- How *confident* do participants feel about detecting attention with and without accessible nonverbal cues?

4.1 Methods

We recruited 16 participants from an organization that provides services to BLV people. The participants identified as blind or low vision (5 females, 11 males) whose ages ranged from 18 to 74 (mean = 43.44, standard deviation (SD) = 44, Table 1).

Each participant took part in a single in-person 90-minute session. One researcher led the study and two researchers participated in the study as conversationalists in the virtual environment. During the study session, participants first completed a tutorial on the VR system and accessible cues. Then, they completed three conversation tasks in virtual environments: (1) a practice task, (2) a baseline task, and (3) a treatment task. The researchers would introduce a new conversation topic at the beginning of each conversation task, including the practice task. We counterbalanced the order of the baseline and treatment tasks. We concluded the session with a semi-structured interview.

4.1.1 VR Prototype. We developed a VR prototype with multi-user virtual environments to give participants a virtual conversation space to test the cues in. These environments allowed participants and researchers acting as conversationalists to join the space and talk together. We designed the prototype to automatically detect when the researchers nodded their heads, shook their heads, or made eye contact with the participant. This automatic detection allowed the researchers to focus on the conversation tasks and their assigned attention conditions (see section 4.1.4) instead of manually triggering cues. Since participants did not need access to their own nonverbal cues, participants' cues were ignored by the system to prevent confusion. The prototype detected head movements

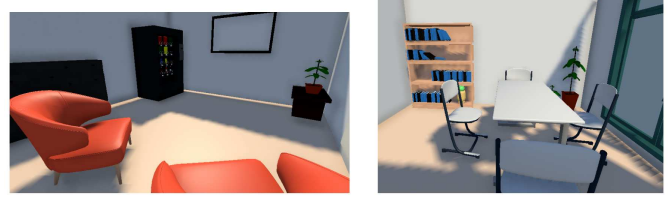


Figure 1: The virtual environments used in the study. Left: the tutorial room, designed to look like a small waiting room. Right: the task room, designed to look like a professional conference room. The same room was used for both baseline and treatment tasks.

by noting repeated translations of the VR headset left and right (shaking head) or up and down (nodding). It detected eye contact using virtual raycasts sent from a “hit box” around the eye levels of each avatar; if these raycasts both reached each other’s hitboxes, the prototype determined eye contact was being made. If any of these behaviors were detected, the prototype triggered the accessible cues, playing the corresponding audio or haptic feedback on the participants’ headset and controllers.

The prototype played all nonverbal cues at once for any behaviors that were detected from the researchers who were in the virtual space. This meant participants often received feedback for accessible cues from the researchers at the same time. To determine who the cues were coming from, participants used a combination of audio spatialization and their own head movements in VR. All audio cues were spatialized so participants heard cues coming from the researcher in the room triggering them. For eye contact, our haptic cue, participants received haptic feedback when they were oriented toward the researcher who was looking at them, i.e., if they felt haptic vibrations when looking left, they knew the researcher on their left was the one making eye contact with them. In this way, participants were easily able to distinguish between different researchers’ cues during conversation.

4.1.2 VR Tutorial. We began the session with a ten-minute VR equipment tutorial and introduction to our accessible nonverbal cues. Then, we explained how to wear a VR headset and use the controllers. We also explained and played each of the three accessible nonverbal cues.

4.1.3 Practice Task. Participants entered the virtual environment and engaged in a ten-minute conversation with two researchers using the cues. The topic was how things have improved after the pandemic. We split the task into two segments, pausing the conversation after five minutes. This allowed us to conduct checks on participants’ thoughts towards the start of the conversation, and towards the end when they were more familiar with their conversation partners. After each segment, we asked three questions:

- Can you describe what you just experienced in the virtual space?
- What were the reactions of the other speakers?
- “What’s your level of confidence in your answer to the previous question on a scale of 1 to 5, where 1 is not at all confident and 5 is very confident?”

Table 1: Participant demographics. F=Female; M=Male; All vision measures (acuity, visual field, light perception) were self-reported by participants.

P#	Age	Gender	Vision	Visual Acuity	Visual Field	Light Perception
P1	37	F	Low vision / Blind	R: 20/320, L: 20/160	R: no peripheral, L: wider visual field	Yes
P2	51	F	Low vision	R: 20/700, L: 20/1200	R: limited peripheral, L: ok	Yes
P3	67	M	Low vision	R: Unsure, L: 20/200	R: Unsure, L: full	Yes, left eye
P4	27	M	Blind	Unsure	Unsure	Yes
P5	21	M	Low vision	Unsure	None	Yes, right eye
P6	18	F	Blind	R: blurry, 2-3 inches, L: No peripheral, 23/200	Unsure	Yes
P7	50	M	Blind	None	None	None
P8	36	M	Low vision / Blind	Unsure	Lacks central, left eye is slightly better	Limited on the peripheral
P9	45	M	Blind	None	None	None
P10	74	F	Blind	None	None	None
P11	26	M	Low vision / Blind	R: 20/300, L: 20/400	Slight central in left eye, 180-degree visual fields	Yes
P12	66	M	R: Low vision, L: Blind	Unsure	Unsure	Yes
P13	34	M	Low vision	Unsure	Unsure	Yes
P14	45	F	Blind	None	None	None
P15	55	M	Low vision	R: Unsure, L: Unsure	Unsure	Yes, left eye
P16	43	M	Blind	None	None	None

Finally, we asked a set of Likert-scale questions and an open-ended question to allow participants to reflect on the experience.

4.1.4 Counterbalanced Baseline and Treatment Tasks. Participants experienced a 15-minute debate task. With this task, we wanted to create a situation where the participant would need to figure out whether attention was on them or not without the researchers explicitly saying anything about their attention state. Thus, we designed the task as a debate since it would require that researchers first pose topics and then fall silent to listen to the participant's response. Two researchers acted as conversation partners for the participant during the debates. The researchers would present a casual topic for debate, such as whether hot dogs were considered a sandwich, and then prompt the participant for their position. Following this, the participants' goal was to convince their conversation partners of their position on the topic. The participants completed this task twice in a counterbalanced order: once with the cues (treatment task) and once without (baseline). They were presented with a new topic for each debate.

Attention Conditions. We created four "attention conditions" as a ground truth of the researchers' attention states (whether the researchers were paying attention to the participant at a given time). Participants were not given these attention conditions; these were for the researchers only. To determine what it meant for a

researcher to be "paying attention," we looked at standard definitions of attention from psychology. One of the most commonly used definitions of attention in prior work comes from James' *Principles of Psychology*, which defines attention as the "[f]ocalization, concentration, of consciousness" on "one out of what seem several simultaneously possible objects or trains of thought" [22]. In other words, "attention" is when a person focuses on one object out of several others. To operationalize this, we had our researchers focus on the participant and *demonstrate* their focus by using verbal and nonverbal cues to respond to whatever the participant was saying in conversation. If they responded to the participants' thoughts and behaviors for over half the duration of the conversation, we considered the researcher to have been focused on the participant for the majority of the conversation. Thus, they were "paying attention" to the participant by standard definitions from psychology.

In total, we had four attention conditions: (1) **both researchers** were paying attention to the participant, (2) **neither researcher** was paying attention to the participant, (3) **only researcher A** was paying attention to the participant, and (4) **only researcher B** was paying attention to the participant.

The 15-minute debate task was split into two 7.5-minute segments in which participants experienced **one** attention condition per segment. Since participants repeated this task twice, once for the baseline and once for the treatment, they experienced **four**

Table 2: A table showing the four different task sets, with two counterbalanced attention conditions and one debate topic per each task.

Set	Debate Topic	Segment 1 Attention Condition	Segment 2 Attention Condition
A	If a hotdog is a sandwich	Nobody pays attention	Only Researcher A pays attention
B	Morning or evening showers are better	Everyone pays attention	Only Researcher B pays attention
C	If pineapple belongs on pizza	Only Researcher B pays attention	Only Researcher A pays attention
D	If sandwiches should have crust or no crust	Everyone pays attention	Nobody pays attention

attention conditions in total. The order in which they experienced the attention conditions was counterbalanced (see table 2). After each segment, we asked the following:

- (1) Can you describe what you just experienced in the virtual space?
- (2) Did **researcher A** pay attention to you?
- (3) Did **researcher B** pay attention to you?
- (4) “What’s your level of confidence in your answer to the previous question on a scale of 1 to 5, where 1 is not at all confident and 5 is very confident?”

This was meant to gauge participants’ confidence level in their ability to determine the attention states of their conversation partners.

At the end of each task, we asked Likert-scale questions to help us understand participants’ experiences with and without cues. We asked participants to rate their agreement with the following statements about the quality of their experience:

- “The accessible nonverbal cues were useful to detect if people were paying attention to me or not”
- “The accessible nonverbal cues were distracting”
- “It was easy to determine whether people were paying attention to me or not”

Additionally, we asked participants to rate their agreement with the following statement: “It was easy for me to determine whether attention was on me or not.” After the treatment task, we asked them to rate their agreement with the following: “The accessible nonverbal cues were distracting” and “The accessible nonverbal cues were useful to detect attention.”

4.1.5 Semi-Structured Interview. We ended the study with a 20-minute semi-structured interview, in which we asked the participants to reflect on their experience and preferences for accessible nonverbal cues. We discussed scenarios where cues may be useful, other nonverbal behaviors, other effects of the cues, and suggestions to improve the cues.

4.1.6 Data and Qualitative Analysis. All participants completed all of the tasks and interviews. Audio and video recordings from the study session were collected and transcribed using an automatic transcription service, Otter.ai. Two researchers coded the transcripts using open descriptive codes. They coded the same two transcripts, then came together and discussed discrepancies. Through discussion, they generated a codebook and split the rest of the data. Afterward, they conducted a thematic analysis [7] using affinity diagrams to categorize the codes into themes.

4.1.7 Statistical Analysis. We also conducted a statistical analysis to identify whether the nonverbal cues had a significant effect on participants’ accuracy and confidence when assessing attention states. We used a linear mixed effect model fitted by Restricted Maximum Likely estimation to model our data. We used this type of model since we had collected repeated measures of participants’ reactions to the conversation at varying segments, as well as to handle cases where participants had refused to answer our measures, creating occasional missing data points. *ParticipantID* was used as a random effect to indicate to our model that multiple measures were attached to each Participant ID (i.e., that we collected multiple measures from our participants). T-tests run on the models used Satterwaite’s method to account for variances in our data samples.

Effect of Nonverbal Cues on Accuracy and Confidence. We explored the effect of the nonverbal cues on participants’ accuracy and confidence in assessing attention states. To do this, we fit one model with one factor, *Group* (two levels: *withCues*, *withoutCues*), one measure for accuracy, *Accuracy*, and one measure for confidence, *Confidence*. The values for *Group* were pulled from the study design, whether users were experiencing the scenario with or without the cues. The *Accuracy* and *Confidence* scores were pulled from participants’ assessments of attention states during each scenario segment. Each participant provided four assessments per task of *their opinion* of whether a particular speaker was paying attention: one for speaker A in segment 1, one for speaker A in segment 2, one for speaker B in segment 1, and one for speaker B in segment 2. We had 16 participants who each provided eight assessments, resulting in 128 total assessments.

4.2 Findings

4.2.1 Task Performance and Statistical Tests. We report participants’ accuracy and confidence when assessing the attention states of their speaking partners. We also report statistical tests to identify whether having the nonverbal cues had a significant effect on participants’ accuracy and confidence. We quantified the usefulness of the cues by examining the impact of one factor, *Group*, with two levels (*withCues* and *withoutCues*) on participants’ performance. We interpreted “significance” as a p-value < 0.05, established prior to running tests.

Accuracy. We report scores for 16 participants (see figure 2). 11 out of 16 participants had better accuracy scores when using the accessible nonverbal cues. Out of the remaining five, four participants scored the same as they did without cues, and one participant scored marginally better (4 correct without cues versus 3 correct with cues).

Accuracy Scores Histogram (WithoutCues Versus WithCues)			Confidence Scores Histogram (WithoutCues Versus WithCues)		
PID	WithoutCues	WithCues	PID	WithoutCues	WithCues
P1	2	3	P1	NA	NA
P2	2	3	P2	NA	NA
P3	2	4	P3	3	5
P4	2	3	P4	4	5
P5	4	4	P5	5	5
P6	3	4	P6	4	4
P7	2	3	P7	5	5
P8	4	4	P8	5	5
P9	4	3	P9	5	3
P10	2	3	P10	5	5
P11	3	4	P11	3	5
P12	2	4	P12	NA	4
P13	3	4	P13	5	5
P14	3	3	P14	NA	3
P15	3	3	P15	5	5
P16	3	4	P16	4	5

Figure 2: Histograms comparing each participants' accuracy and confidence scores between the task without cues and the task with cues. Red indicates worse scores with cues, yellow indicates equal scores, and green indicates better scores with cues. Gray indicates one or more of the answers were not provided by participants (NA) so a comparison cannot be made.

Effect of Nonverbal Cues on Accuracy. We found a significant effect of *withCues* on *Accuracy* ($t=2.614$, $p=0.010$). The mean accuracy of participants across all conversation scenarios while using the cues was 87.5% ($SE=0.1$), compared to 68.8% accuracy ($SE=0.1$) without the cues. We conclude that participants' accuracy was significantly higher when using the cues. Thus, the cues helped their ability to accurately judge how much attention people were paying to them during conversation.

Confidence. We report scores for 12 participants (see figure 2). The remaining four did not provide confidence scores. Half the participants, six out of 12, were more confident when using the accessible nonverbal cues. Out of the remaining six, three of them scored their confidence the same as they did without cues, and the other three reported higher confidence without cues. However, these three participants all tended to score their confidence very high across all conditions, having an average score of 17.7 for confidence out of a maximum confidence score of 20 across all conditions.

Effect of Nonverbal Cues on Confidence. We found a significant effect of *withCues* on *Confidence* ($t=2.224$, $p=0.028$). The mean confidence participants reported across all conversation scenarios while using the cues was 4.0 ($SE=0.4$) out of 5, compared to a confidence score of 3.7 ($SE=0.4$) out of 5 without the cues. We conclude that participants' confidence was significantly higher when using the cues. Thus, the cues improved how confident participants felt that they knew how much attention people were paying to them, and so they were more willing to act on their judgments of attention.

Quality of Experience. Participants provided scores on a scale of 1 to 5, where 1 meant they strongly disagreed and 5 they strongly

Usefulness and Distraction Likert Scale Responses Histogram						
1	2	3	4	5	Mean	SD
The accessible nonverbal cues were useful to detect if people were paying attention to me or not						
1	2	3	2	8	3.87	1.32
The accessible nonverbal cues were distracting						
3	6	3	3	1	2.53	1.2

Figure 3: Likert-scale score responses for cue usefulness and distraction.

agreed with a given statement about the quality of their experience with the cues. Each statement, also reported in section 4.1.4, related to a quality metric as follows:

- [Usefulness] “The accessible nonverbal cues were useful to detect if people were paying attention to me or not”
- [Distraction] “The accessible nonverbal cues were distracting”
- [Task Ease] “It was easy to determine whether people were paying attention to me or not”

Usefulness. Scoring usefulness helped us understand whether participants found the cues useful, regardless of other difficulties with the cues. Participants reported high scores for usefulness (see figure 3). 10 out of 16 participants agreed or strongly agreed the nonverbal cues were useful, and three neither agreed nor disagreed. Interestingly, the three participants who disagreed or strongly disagreed had low vision. These results indicate that most participants found the cues useful, but low vision people might find them less useful than blind people (possibly due to their residual vision).

Distraction. Scoring distraction helped us understand whether participants felt distracted from their conversation with the other speakers when using cues. Participants reported relatively low scores for distraction (see figure 3). Nine out of 16 participants disagreed or strongly disagreed, and three participants neither agreed nor disagreed. The four participants who agreed or strongly agreed had low vision, as seen above with usefulness. These results indicate that most participants did not find the nonverbal cues distracting, but that low vision people might find them more distracting than blind people (possibly due to having to split their attention between their residual vision, the cues, and the speakers' speech).

Task Ease. We compared the metric of task ease between the *withCues* and *withoutCues* conditions. This helped us understand how much effort participants felt they had to put in to determine attention states. Participants generally reported higher scores for task ease with cues (see figure 4). Nine out of 16 participants reported higher ease scores with cues than without cues, three reported the same scores with and without cues, and four reported lower scores with the cues. These results indicate that most participants felt it took less effort to determine attention states when the cues were present, though a small minority found it more challenging.

4.2.2 Participants' Reactions to the Cues. We report participants' reactions to the cues in terms of their usefulness, usability, and emotional impact, three metrics often used when evaluating user experience [3, 37].

Usefulness. Most participants felt the cues were useful, allowing them to determine others' behaviors in ways they normally could

Task Ease Likert Scale Responses Histogram (<i>WithCues</i> Versus <i>WithoutCues</i>)						
1	2	3	4	5	Mean	SD
It was easy to determine whether people were paying attention to me or not (<i>WithCues</i>)					4.18	0.8
0	1	1	8	6		
It was easy to determine whether people were paying attention to me or not (<i>WithoutCues</i>)					3.62	0.93
0	3	2	9	2		

Figure 4: Likert-scale score responses for task ease with cues and without cues.

not. For example, the cues helped P9 determine if people were ignoring him in the conversation: “I feel like there was less attention being paid [to me]...I was hearing fewer of the head nods and shakes and making eye contact was a little more difficult, too.” P11 used the cues to reassure himself that others had been paying attention to him: “They were...engaging with my points that I brought up and nodding or not nodding towards my answers.” Overall, we found that eight participants used the cues to determine attention by connecting the amount of nonverbal cues they noted with the level of attention given to them.

While many participants found the cues useful, others were frustrated by them. Three participants felt they were able to converse adequately without the cues, and found the addition of cues “distracting.” These participants mentioned the cues might be useful in certain scenarios, such as when no one is offering them verbal feedback. However, since they did not experience that in the study, the cues did not feel necessary: “There’s always one person to talk to. Maybe if none of them were talking to me then it might be a little different.” (P14)

Participants also sometimes disagreed with the information the cues provided about the speakers, finding that it conflicted with the information they had determined themselves. We especially noted this among our participants with residual vision, who could see movements from the speakers’ avatars (P15, P4). For example, P4 mentioned he “could see the head turning” when speaking with the other avatars, as well as their basic gestures. As such, he preferred to rely on “body language and [head movements]” that he could see, instead of “believing” (P4) the nonverbal cues.

Finally, a few participants doubted whether the cues were useful outside of the study. One participant mentioned he would not use cues in real life until he knew a person well enough to understand “what their movements mean” and thus decide which cues he should be looking for (P15). Another participant, P1, did not want to receive cues from people when he wasn’t conversing with them:

I certainly wouldn’t want to wear this in, like, outside settings, like riding on the bus. Apparently, people look at me all the time. Whenever I am on the bus my family tells me, ‘Oh, that person’s looking at you,’ and I don’t care. I’m just doing what I need to do. [So] this wouldn’t be useful for me, like, I’m seeing everybody stare at me that’d make me so self-conscious. I don’t know how you guys walk around the world knowing everybody’s looking at you. It’s kind of freeing to not know that. (P1)

Usability. Participants gave varying feedback on whether they found the cues easy to learn and use. Twelve participants felt the learning curve was low, and commented that the cues seemed like a natural way to provide information about nonverbal behaviors. P9 said, “You get used to [the cues] pretty quickly because it’s just a pretty natural thing, and the sounds really are not overpowering. So they would just become part of the scenery, the...auditory scenery.” Two participants agreed that the cues became part of the conversation’s “background” and felt they did not have to “pay attention that much” (P16) to recall the cues’ meaning and apply it to conversation.

Four participants mentioned that it was difficult to remember what the cues signified. For instance, P15 mentioned he could not recall what the cues meant since he would end up being more worried “about how I look to someone” because of the constant eye gaze feedback. P1 also expressed her frustration with trying to connect what the cues were signaling to the speakers’ behaviors. She explained that “The haptic buzzes [don’t help with] figuring out who’s looking at me, both of these [controllers] were just buzzing constantly. So I’m assuming everybody was looking at me” (P1).

Emotional Impact. While most participants did not have strong emotional reactions to the cues, some were moved by how much the cues impacted their confidence during conversation. P2 in particular felt empowered by having the cues: “I felt very confident...because in my real [life], I don’t get to understand head nods and eye contact...It was empowering to get that feedback.” She explained that this sense of empowerment was heightened by being able to respond to the feedback she was getting, giving her a sense of control. “I really like the autonomy and the ability to know when you’re agreeing or disagreeing, instead of having to wait for someone to tell me...there’s something kind of powerful about actually knowing.” P5 agreed with these sentiments, explaining that for him, having the cues “reassures that they’re listening to me” and gives him the confidence to keep speaking.

4.2.3 Uses of Cues. We now report participants’ reflections on possible uses for the cues.

Scenarios. Our participants gave a wide range of scenarios that they imagined cues could be useful for. These scenarios included: Intimate or Private Interactions:

- Going on first dates (P4)
- Watching shows with friends or family (P14)

Professional Interactions:

- Attending virtual conferences (P7)
- Attending or facilitating meetings at work (P7)
- Interviewing for a job (P4)
- Having a one-on-one conversation with your boss (P16)

One of our participants, P6, also brought up a social scenario she encounters frequently in her daily life, where sighted people either assume she can see or forget about her visual impairment. “There are those people that we come across who can forget that we can’t see, and they’ll just shake their head. Yes or no.” In these scenarios, P6 envisioned the cues could provide clarity by telling her about nonverbal behaviors, and prevent sighted peoples’ misconceptions “that we’re ignoring you” by helping her respond to those behaviors on time.

Finally, two of our participants (P16 and P2), who were both educators, felt the cues could be helpful for interactions with students. In P16's case especially, he felt the cues would help him meet access needs for his students:

I'm a special education teacher and some of the students I work with are non-verbal. So they have developmental disabilities. They don't speak. And all they do is gesture or point at something if they want something, and then in that case I don't see what they're looking at or what they're pointing at. So they work with, in the classroom, they have aids... So I ask them, 'Is he looking at me, or is he looking at something else? Is he looking at the toy?... Is he nodding [or] doing a different gesture?' So if I can do that with smart glasses or something similar like this virtual headset [with cues]...I think it will help me with my work. (P16)

Being A Better Conversationalist. Some participants felt the cues made them better conversationalists, and "leveled the playing field" (P2) when speaking with sighted people so they could participate in more ways than usual. For example, P2 felt that the cues allowed her to take a leadership role and guide conversation topics.

It was rather empowering to be in [the virtual conversation] and to be able to say, 'Oh, so you agree, or so you disagree. Tell me more about that.' I feel like [the cues] almost gave me a stronger position of leadership where I can decide where to go next. (P2)

Other participants mentioned they felt the cues improved the flow of the conversation. P13 explained that often, people only give him "a head nod or a shake. And I'm just like, 'Okay, I'll just stare at you until an answer comes, or you ask me... what I'm doing.'" He felt the cues removed that step of waiting awkwardly for clarification. As a result, "the repartee [of the conversation felt] a lot smoother because I can say something that's kind of an open-ended question, and just get a nod or a no, and then continue going instead of having the conversation stop and have the other person have to say, yes or no."

4.2.4 Design Suggestions. We now discuss various suggestions participants shared to improve the nonverbal cues.

Modality Suggestions. Participants had various preferences for the modality (haptic or audio) of the cues. Several participants felt that haptics were less disruptive than audio, commenting that haptics are not "overpowering the conversation or overpowering whoever is talking" (P8) and they "wouldn't distract your engagement to the conversation" (P10). P2 also added, "I like the haptic because it leaves room for my listening to be focused on everything else and what people are saying."

However, some participants mentioned that haptics may not be as versatile as audio, where you can use various sounds to differentiate cues. P16 noted that haptics could be difficult to use for multiple nonverbal cues since "we are more limited" to simple vibrations which "might be hard to differentiate" from each other, especially if the system is trying to signal more than one cue at once. To meet this shortcoming, some participants suggested using both haptics

and audio to have "multiple ways of disseminating information" (P2) during a conversation.

P9 explained that modality preferences could vary based on user ability and suggested that both haptics and audio should remain to account for this. P9 explained that for "someone who's deaf-blind, having that haptic feedback is a big one that could help them be more engaged in communication. ... But, on the other hand, if you have someone who's lost sensitivity in their hands, haptic is not great, either. So I say, keep both paths available" (P9).

Suggestions for Additional Cues. Participants had various suggestions for additional cues. Some requested more gestures including someone looking at their cell phone (P14), raising their hand (P7), or reaching out for a handshake (P13).

Participants also wanted information about more subtle nonverbal behaviors. P1 was interested in receiving specialized eye gaze information, such as having one cue for concentrated eye contact versus another for a quick glance. P7 said that facial expressions like a "dirty look" or a "look of total disengagement" would be helpful. While most participants were interested in more cues, others brought up concerns that subtle cues may occur too often: "There just seems to be so much more possibilities... that I wouldn't want to get information for, like slightly sad or slightly happy. It seems like an information overload." (P11)

5 DISCUSSION

Addressing our research question, we found our designs of accessible nonverbal cues for eye contact, head nodding, and head shaking effectively supported conversations in VR for BLV users. Moreover, our designs provided participants with a concrete experience of holding real conversations in VR with access to nonverbal behaviors, which allowed them to reflect on possibilities for accessing more forms of nonverbal social information. The design and evaluation of the nonverbal cues allowed us to identify novel design implications for nonverbal cues.

In the following sections, we (1) discuss how our findings relate to prior work on the design of nonverbal cues in virtual environments, (2) introduce cue banks, an approach to address the challenge of overwhelming the users with too many types of nonverbal cues, and (3) discuss considerations for future AI-powered recognition of nonverbal cues.

Another Step Towards VR Accessibility. Prior work has explored nonverbal behaviors that are commonly used in social VR, such as head movements [1, 31] and hand gestures [31, 45]. Prior work has also emphasized the importance of nonverbal behaviors in interactions between users, such as that nodding between users and avatars resulted in higher levels of trust towards avatars [1]. Our designs of accessible nonverbal cues add to this body of work to allow BLV users to receive information about nonverbal behaviors that are commonly used in a virtual environment.

In prior work, Wieland et al. [49] explored different modalities to represent eye gaze in VR and found that participants had positive responses to using audio and haptic to represent eye gaze. Our evaluation adds to Wieland et al.'s findings, showing that audio and haptic designs can be effective ways of representing nonverbal behaviors, but also that there is great variability in how people respond to these modalities in practice. For instance, nearly all

participants could easily perceive haptics without being distracted from conversation. However, continuous vibrations can be overwhelming or overstimulating, even if they are easy to maintain in the background of a conversation.

In another study, Wieland et al. [48] identified which nonverbal behaviors are important to incorporate into conversations in VR, and suggested that multiple kinds of cues should be incorporated into conversation. While our work echoed this suggestion, and our participants suggested a variety of new cues to incorporate into future systems, we also found that including too many cues at once becomes quickly overwhelming in practice. Therefore, new methods of handling large sets of cues should be considered.

These suggestions are centered on making social VR accessible by supporting communication for BLV users. However, there are many other directions for future work to address social VR inaccessibility. For instance, while cue systems can help solve communication challenges in one-on-one conversations or small groups, future work can consider how these systems could support larger social venues in VR, such as a music concert [6]. Researchers should consider whether cue systems could provide information about large and complex groups without detracting from a BLV user's enjoyment and immersion in these scenarios.

Cue Banks. Both our participants and prior work suggested several cues to add to future systems. However, if all of these cues were implemented and provided to users at once, they would likely become overwhelming. Future work could explore options for handling large sets of cues in ways that will not overwhelm BLV users.

One way to address this would be “cue banks.” These can contain all possible nonverbal cues that the system can represent. Within these larger banks, there can be sub-banks of cues that are commonly used in different situations. Users could select and activate specific sub-banks to receive smaller sets of cues that they find relevant to their social situations. We have developed some example sub-banks which include cues that our participants suggested (see section 4.2.4), grouped with other possible cues we believe fall under similar categories of use:

- Boredom cues: looking at a mobile device, glazed-over eyes, nodding off
- Greeting cues: waving, holding out hand for handshake
- Listening cues: eye gaze, head nodding, head shaking
- Reaction cues: scowl, wide eyes of surprise, big smile
- Subtle eye cues: judgemental stare, suspicious stare, interested stare
- Facilitating cues: raising hands, confused expression, “boredom cues”

As seen above, sub-banks would vary in size, such as the “facilitating cues” sub-bank which may contain smaller sub-banks like “boredom cues.” Future work could investigate the ideal sizes of sub-banks, or additional kinds of sub-banks that BLV users may find useful. Another approach the research community could examine is allowing users to apply sub-banks to specific people in specific social scenarios. For example, if a BLV user is giving a presentation to their coworkers and boss, they may be looking for different social information among the audience members. They might assign the “reaction cues” to their boss, so they know if their boss likes the

ideas they are presenting. For everyone else, they might only attach the basic “listening cues” to ensure they are paying attention.

It is important to also consider how these sub-banks might be integrated with existing commercial VR platforms, such as VRChat and RecRoom. These existing platforms often include chaotic, uncontrolled environments, where such sub-banks may introduce challenges such as information overload due to the large number of users within the environment.

Finally, rather than relying on the users themselves to assign and activate different sub-banks, future cue systems could attempt to “read the room,” and activate certain sub-banks based on the room's liveliness or atmosphere. For example, if the room becomes unusually silent for an extended period of time, the cue system can automatically turn on “boredom cues” or “listening cues” to help the user assess the situation. If the room becomes lively again, the system could turn off the previous sub-bank and apply “reaction cues” instead. Supposing a new speaker approached the BLV user now that the mood lifted, the system could quickly turn on “greeting cues,” ensuring the user notices when the new speaker holds out a hand to introduce themselves.

While the above system designs have great potential, it is important to first acknowledge that there are cultural differences surrounding nonverbal cues. These differences may make the process of assigning sets of cues with specific meanings (such as “boredom” or “greeting” cues) much less certain. This study was conducted in the United States, where nonverbal cues may have very different meanings compared to how they are used in other countries. For example, while eye contact in the United States is commonly understood to represent someone paying attention, in other cultures such as some “Asian, African, and Latin American countries, eye contact should be avoided to show respect” [46]. Further, there may be cultural differences in social cues used by sighted and BLV people. While our work gave BLV people access to sighted-normative cues to support conversations, in a situation where a sighted person was in a room full of blind people, there might be BLV-normative cues of an entirely different kind that the sighted person would need to carry conversation. Future work should take cultural differences into account when designing and evaluating nonverbal cues in different countries or mixed-ability spaces.

Challenges with AI-Powered Detection of Nonverbal Behaviors. Another important consideration for future cue systems is implementing AI in accessible VR, as the use of AI in VR is growing rapidly. Many of the challenges of implementing AI in VR are well-documented [9, 29, 40], so we focus our discussion here on the unique difficulties and opportunities we foresee with using AI for nonverbal behaviors. First, AI-based detection introduces challenges due to its inaccuracy with identifying many facial expressions and gestures. Current face recognition technologies are known to have difficulties determining even basic emotional expressions such as happiness or sadness with high accuracy, let alone subtle differences between types of eye gaze [5]. One potential way to address this uncertainty may be to modify the nonverbal cue feedback. For example, when a user receives audio feedback that a conversational partner smiled, if the AI has 99% certainty the event happened, the audio feedback could be higher in volume. In contrast, if the AI has a lower level of certainty, the audio feedback could be lower in volume. Haptic versions of these feedback systems

could be applied as well, so that participants can receive confidence information via one channel (tactile feedback from haptics) while they receive the accessible cue itself via another (audio feedback from the accessible cue). This would allow users to quickly assess the certainty of the detection system as they converse.

In addition, there are possible privacy concerns with AI constantly tracking people's nonverbal behaviors in order to implement such a system. Some users may feel uncomfortable with AI collecting such information, particularly if they are not made aware that the system is observing and translating their behavior for another user. One possible solution would be for other users to provide "consent" to share their nonverbal behaviors when a system tries to access them, although this could inadvertently expose the BLV user's visual condition by telling users there is a person with a visual impairment nearby. Another possibility is to have the system only detect the behaviors of those on the BLV user's friend list, or make consent to share nonverbal behaviors a part of the friend request system. In general, future work should take privacy concerns into account when creating systems that track nonverbal behaviors.

6 CONCLUSION

In this paper, we designed and evaluated accessible audio and haptic cue representations of three nonverbal behaviors: eye contact, head shaking, and head nodding. Our evaluation of the accessible nonverbal cues with 16 blind and low vision participants highlighted that the designs of the accessible nonverbal cues were effective in supporting conversations. We also uncovered a range of participant preferences for current cue designs and various approaches the research community could examine for future cue designs. Our work points to a novel avenue of making social VR fully accessible for blind and low vision people.

REFERENCES

- [1] Nadine Aburumman, Marco Gillies, Jamie A Ward, and Antonia F de C Hamilton. 2022. Nonverbal communication in virtual reality: Nodding as a social signal in virtual interactions. *International Journal of Human-Computer Studies* 164 (2022), 102819. <https://www.sciencedirect.com/science/article/pii/S1071581922000489#ecom0001>
- [2] Ronny Andrade, Steven Baker, Jenny Waycott, and Frank Vetere. 2018. Echo-house: Exploring a virtual environment by using echolocation. In *Proceedings of the 30th Australian conference on computer-human interaction*. 278–289. <https://dl.acm.org/doi/abs/10.1145/3292147.3292163>
- [3] LinkedIn Community Articles. n.d.. How do you evaluate the emotional impact of your UX design? <https://www.linkedin.com/advice/0/how-do-you-evaluate-emotional-impact-your-ux-design>
- [4] Jeremy N Bailenson, Andrew C Beall, Jack Loomis, Jim Blascovich, and Matthew Turk. 2004. Transformed social interaction: Decoupling representation from behavior and form in collaborative virtual environments. *Presence: Teleoperators & Virtual Environments* 13, 4 (2004), 428–441. <https://direct.mit.edu/pvar/article-abstract/13/4/428/18516/Transformed-Social-Interaction-Decoupling>
- [5] Andrew Begel, John Tang, Sean Andrist, Michael Barnett, Tony Carbary, Piali Choudhury, Edward Cutrell, Alberto Fung, Sasa Junuzovic, Daniel McDuff, et al. 2020. Lessons learned in designing ai for autistic adults. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–6. <https://andrewbegel.com/papers/lessons-learned-autism.pdf>
- [6] Viverse Blog. n.d.. Exploring the VR Music Revolution: Top 5 VRChat Music Concerts, Videos, and Performances. <https://www.news.viverse.com/post/exploring-the-vr-music-revolution-top-5-vrchat-music-concerts-and-performances>
- [7] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101. <https://www.tandfonline.com/doi/abs/10.1191/1478088706qp0630a>
- [8] Global Listening Centre. n.d.. Body Language of Listeners. <https://www.globallisteningcentre.org/body-language-of-listeners>
- [9] Pietro Cipresso and Giuseppe Riva. 2015. Virtual Reality for Artificial Intelligence: human-centered simulation for social science. *Annual Review of Cybertherapy and Telemedicine 2015: Virtual Reality in Healthcare: Medical Simulation and Experiential Interface* (2015). https://books.google.com/books?hl=en&lr=&id=_MbXCwAAQBAJ&oi=fnd&pg=PA177&dq=artificial+intelligence+and+virtual+reality&ots=d49mPiIo1v&sig=BXFBmLmz1yIQLQ1erc1PK3vBYv=onepage&q=artificial%20intelligence%20and%20virtual%20reality&f=false
- [10] Jazmin Collins, Crescentia Jung, and Shiri Azenkot. 2023. Making Avatar Gaze Accessible for Blind and Low Vision People in Virtual Reality: Preliminary Insights. In *2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 701–705. <https://ieeexplore.ieee.org/abstract/document/10322247>
- [11] Jazmin Collins, Crescentia Jung, Yeonju Jang, Danielle Montour, Andrea Stevenson Won, and Shiri Azenkot. 2023. "The Guide Has Your Back": Exploring How Sighted Guides Can Enhance Accessibility in Social Virtual Reality for Blind and Low Vision People. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–14. <https://dl.acm.org/doi/abs/10.1145/3597638.3608386>
- [12] Loredana Crisan. 2021. Emojis finally have a voice: Introducing Soundmojis on Messenger. <https://messengernews.fb.com/2021/07/15/emojis-finally-have-a-voice-introducing-soundmojis-on-messenger/>
- [13] Aaron Gluck, Kwajo Boateng, and Julian Brinkley. 2021. Racing in the dark: Exploring accessible virtual reality by developing a racing game for people who are blind. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 65. SAGE Publications Sage CA: Los Angeles, CA, 1114–1118. <https://journals.sagepub.com/doi/abs/10.1177/1071181321651224>
- [14] Aaron Gluck and Julian Brinkley. 2020. Implementing "The Enclosing Dark": A VR Auditory Adventure. *Journal on Technology and Persons with Disabilities* 8, 2020 (2020), 149–159. <https://scholarworks.csun.edu/bitstream/handle/10211.3/215985/2197%20Implementing%20The%20Enclosing%20Dark%20A%20VR%20Auditory%20Adventure.pdf>
- [15] Inês Gonçalves, André Rodrigues, Tiago Guerreiro, and João Guerreiro. 2023. Inclusive Social Virtual Environments: Exploring the Acceptability of Different Navigation and Awareness Techniques. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–7.
- [16] Ricardo E Gonzalez Penuela, Wren Poremba, Christina Trice, and Shiri Azenkot. 2022. Hands-On: Using Gestures to Control Descriptions of a Virtual Environment for People with Visual Impairments. In *Adjunct Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*. 1–4. <https://dl.acm.org/doi/abs/10.1145/3526114.3558669>
- [17] Ria J Gualano, Lucy Jiang, Kexin Zhang, Andrea Stevenson Won, and Shiri Azenkot. 2023. "Invisible Illness Is No Longer Invisible": Making Social VR Avatars More Inclusive for Invisible Disability Representation. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–4. <https://dl.acm.org/doi/abs/10.1145/3597638.3614480>
- [18] João Guerreiro, Yujin Kim, Rodrigo Nogueira, SeungA Chung, André Rodrigues, and Uran Oh. 2023. The design space of the auditory representation of objects and their behaviours in virtual reality for blind people. *IEEE Transactions on Visualization and Computer Graphics* 29, 5 (2023), 2763–2773. <https://ieeexplore.ieee.org/abstract/document/10049631>
- [19] Yu Hao, Junchi Feng, John-Ross Rizzo, Yao Wang, and Yi Fang. 2022. Detect and approach: Close-range navigation support for people with blindness and low vision. In *European Conference on Computer Vision*. Springer, 607–622. https://link.springer.com/chapter/10.1007/978-3-031-25075-0_41
- [20] Daniel Hung, Heji Kim, and Bruce Land. 1996. Modeling Six Universal Emotions. <https://people.ece.cornell.edu/land/OldStudentProjects/cs490-95to96/HJKIM/emotions.html>
- [21] Masahiro Ide, Shoji Oshima, Shingo Mori, Masato Yoshimi, Junko Ichino, and Shunichi Tano. 2020. Effects of avatar's symbolic gesture in virtual reality brainstorming. In *Proceedings of the 32nd Australian conference on human-computer interaction*. 170–177. <https://dl.acm.org/doi/abs/10.1145/3441000.3441081>
- [22] William James. 2007. *The principles of psychology*. Vol. 1. Cosimo, Inc. <https://infomotions.com/sandbox/great-books-redux/corpus/html/principles.html>
- [23] Tiger F Ji, Brianna Cochran, and Yuhang Zhao. 2022. Vrbubble: Enhancing peripheral awareness of avatars for people with visual impairments in social virtual reality. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–17. <https://dl.acm.org/doi/abs/10.1145/3517428.3544821>
- [24] Stevanus Kevin, Yun Suen Pai, and Kai Kunze. 2018. Virtual gaze: exploring use of gaze as rich interaction method with virtual agent in interactive virtual reality content. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*. 1–2. <https://dl.acm.org/doi/abs/10.1145/3281505.3281587>
- [25] Diana Kornbrot, Paul Penn, Helen Petrie, Stephen Furner, and Andrew Hardwick. 2007. Roughness perception in haptic virtual reality for sighted and blind people. *Perception & psychophysics* 69 (2007), 502–512. <https://link.springer.com/article/10.3758/bf03193907>

- [26] Marco Kurzweg, Jens Reinhardt, Wladimir Nabok, and Katrin Wolf. 2021. Using body language of avatars in vr meetings as communication status cue. In *Proceedings of Mensch und Computer 2021*. 366–377. <https://dl.acm.org/doi/abs/10.1145/3473856.3473865>
- [27] Rebecca Lawson. 2015. I just love the attention: implicit preference for direct eye contact. *Visual Cognition* 23, 4 (2015), 450–488. <https://www.tandfonline.com/doi/abs/10.1080/13506285.2015.1039101>
- [28] Joan Llobera, Bernhard Spanlang, Giulio Ruffini, and Mel Slater. 2010. Proxemics with multiple dynamic characters in an immersive virtual environment. *ACM Transactions on Applied Perception (TAP)* 8, 1 (2010), 1–12. <https://dl.acm.org/doi/abs/10.1145/1857893.1857896>
- [29] Michael Luck and Ruth Aylett. 2000. Applying artificial intelligence to virtual reality: Intelligent virtual environments. *Applied artificial intelligence* 14, 1 (2000), 3–32. <https://www.tandfonline.com/doi/abs/10.1080/088395100117142>
- [30] Maruricio Lumbieras and Jaime Sánchez. 1999. Interactive 3D sound hyperstories for blind children. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 318–325. <https://dl.acm.org/doi/abs/10.1145/302979.303101>
- [31] Divine Maloney, Guo Freeman, and Donghee Yvette Wohn. 2020. "Talking without a Voice" Understanding Non-verbal Communication in Social Virtual Reality. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–25. <https://dl.acm.org/doi/abs/10.1145/3415246>
- [32] Changing Minds. n.d. Attentive Body Language. https://changingminds.org/techniques/body/attentive_body.htm
- [33] Tony Morelli, John Foley, Luis Columma, Lauren Lieberman, and Eelke Folmer. 2010. VI-Tennis: a vibrotactile/audio exergame for players who are visually impaired. In *Proceedings of the Fifth International Conference on the Foundations of Digital Games*. 147–154. <https://dl.acm.org/doi/abs/10.1145/1822348.1822368>
- [34] Vishnu Nair, Jay L Karp, Samuel Silverman, Mohar Kalra, Hollis Lehv, Faizan Jamil, and Brian A Smith. 2021. Navstick: Making video games blind-accessible via the ability to look around. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. 538–551. <https://dl.acm.org/doi/abs/10.1145/3472749.3474768>
- [35] Vishnu Nair, Shao-en Ma, Ricardo E. Gonzalez Penuela, Yicheng He, Karen Lin, Mason Hayes, Hannah Huddleston, Matthew Donnelly, and Brian A. Smith. 2022. Uncovering Visually Impaired Gamers' Preferences for Spatial Awareness Tools Within Video Games (ASSETS '22). Association for Computing Machinery, New York, NY, USA, Article 6, 16 pages. <https://doi.org/10.1145/3517428.3544802>
- [36] Georgios Nikolakis, Dimitrios Tzovaras, Serafim Moustakidis, and Michael G Strintzis. 2004. Cybergrasp and phantom integration: Enhanced haptic access for visually impaired users. In *9th Conference Speech and Computer*. https://www.isca-archive.org/specom_2004/nikolakis04_specom.pdf
- [37] Cruose Okwong. 2023. Usability and usefulness as user experience best practices. <https://bootcamp.uxdesign.cc/usability-and-usefulness-as-user-experience-best-practices-93a43f34f9ac>
- [38] Sébastien Paquette, Isabelle Peretz, and Pascal Belin. 2013. The "Musical Emotional Bursts": a validated set of musical affect bursts to investigate auditory affective processing. *Frontiers in psychology* 4 (2013), 52505. <https://www.frontiersin.org/articles/10.3389/fpsyg.2013.00509/full>
- [39] Christine E Parsons, Katherine S Young, Michelle G Craske, Alan L Stein, and Morten L Kringelbach. 2014. Introducing the Oxford Vocal (OxVoc) Sounds database: a validated set of non-acted affective sounds from human infants, adults, and domestic animals. *Frontiers in psychology* 5 (2014), 562. <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2014.00562/full>
- [40] Tainá Ribeiro de Oliveira, Brenda Biancardi Rodrigues, Matheus Moura da Silva, Rafael Antonio N. Spinassé, Gabriel Giesen Ludke, Mateus Ruy Soares Gaudio, Guilherme Iglesias Rocha Gomes, Luan Guio Cotini, Daniel da Silva Vargens, Marcelo Queiroz Schimidt, et al. 2023. Virtual reality solutions employing artificial intelligence methods: A systematic literature review. *Comput. Surveys* 55, 10 (2023), 1–29. <https://dl.acm.org/doi/full/10.1145/3565020>
- [41] Jonathan Segal, Samuel Rodriguez, Akshaya Raghavan, Heysil Baez, Crescentia Jung, Jazmin Collins, Shiri Azenkot, and Andrea Stevenson Won. 2024. Social-CueSwitch: Towards Customizable Accessibility by Representing Social Cues in Multiple Senses. In *To appear in Proceedings of the 26th International ACM Conference on Human Factors in Computing Systems*.
- [42] Mike Sinclair, Eyal Ofek, Mar Gonzalez-Franco, and Christian Holz. 2019. Capstan-crunch: A haptic vr controller with user-supplied force feedback. In *Proceedings of the 32nd annual ACM symposium on user interface software and technology*. 815–829. <https://dl.acm.org/doi/abs/10.1145/3332165.3347891>
- [43] Alexa F Siu, Mike Sinclair, Robert Kovacs, Eyal Ofek, Christian Holz, and Edward Cutrell. 2020. Virtual reality without vision: A haptic and auditory white cane to navigate complex virtual worlds. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–13. <https://dl.acm.org/doi/abs/10.1145/3313831.3376353>
- [44] Autumn Sprabary. 2022. How eye contact can help or hurt communication. <https://www.allaboutvision.com/resources/human-interest/importance-of-eye-contact/>
- [45] Theresa Jean Tanenbaum, Nazley Hartoonian, and Jeffrey Bryan. 2020. "How do I make this thing smile?" An Inventory of Expressive Nonverbal Communication in Commercial Social Virtual Reality Platforms. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13. <https://dl.acm.org/doi/pdf/10.1145/3313831.3376606>
- [46] Charles Tidwell. n.d. Non-Verbal Communication Modes. <https://www.andrews.edu/~tidwell/bsad560/NonVerbal.html>
- [47] Ryan Wedoff, Lindsay Ball, Amelia Wang, Yi Xuan Khoo, Lauren Lieberman, and Kyle Rector. 2019. Virtual showdown: An accessible virtual reality game with scaffolds for youth with visual impairments. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–15. <https://dl.acm.org/doi/abs/10.1145/3290605.3300371>
- [48] Markus Wieland and Tonja Machulla. 2022. Towards Inclusive Conversations in Virtual Reality for People with Visual Impairments. (2022). <https://dl.gi.de/server/api/core/bitstreams/54742add-e79b-458b-8cd2-6cdbf40c4132/content>
- [49] Markus Wieland, Michael Sedlmair, and Tonja-Katrin Machulla. 2023. Vr, gaze, and visual impairment: An exploratory study of the perception of eye contact across different sensory modalities for people with visual impairments in virtual reality. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–6. <https://visvar.github.io/pdf/wieland2023vr.pdf>
- [50] Markus Wieland, Lauren Thevin, Albrecht Schmidt, and Tonja Machulla. 2022. Non-verbal communication and joint attention between people with and without visual impairments: Deriving guidelines for inclusive conversations in virtual realities. In *International Conference on Computers Helping People with Special Needs*. Springer, 295–304. https://link.springer.com/chapter/10.1007/978-3-031-08648-9_34
- [51] Nick Yee, Jeremy N Bailenson, Mark Urbanek, Francis Chang, and Dan Merget. 2007. The unbearable likeness of being digital: The persistence of nonverbal social norms in online virtual environments. *CyberPsychology & Behavior* 10, 1 (2007), 115–121. <https://www.liebertpub.com/doi/abs/10.1089/cpb.2006.9984>
- [52] Kexin Zhang, Elmira Deldari, Zhicong Lu, Yaxing Yao, and Yuhang Zhao. 2022. "it's just part of me:" understanding avatar diversity and self-presentation of people with disabilities in social virtual reality. In *Proceedings of the 24th international ACM SIGACCESS conference on computers and accessibility*. 1–16. <https://dl.acm.org/doi/abs/10.1145/3517428.3544829>
- [53] Lei Zhang, Klewin Wu, Bin Yang, Hao Tang, and Zhigang Zhu. 2020. Exploring virtual environments by visually impaired using a mixed reality cane without visual feedback. In *2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 51–56. <https://ieeexplore.ieee.org/abstract/document/9288373>
- [54] Yuhang Zhao, Cynthia L Bennett, Hrvoje Benko, Edward Cutrell, Christian Holz, Meredith Ringel Morris, and Mike Sinclair. 2018. 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*. 1–14. <https://dl.acm.org/doi/abs/10.1145/3173574.3173690>
- [55] Yuhang Zhao, Edward Cutrell, Christian Holz, Meredith Ringel Morris, Eyal Ofek, and Andrew D Wilson. 2019. 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*. 1–14. <https://dl.acm.org/doi/abs/10.1145/3290605.3300341>