Hearing the Smile: Scoring Spontaneous Social Scenes for the Visually Impaired

ANONYMOUS AUTHOR(S)

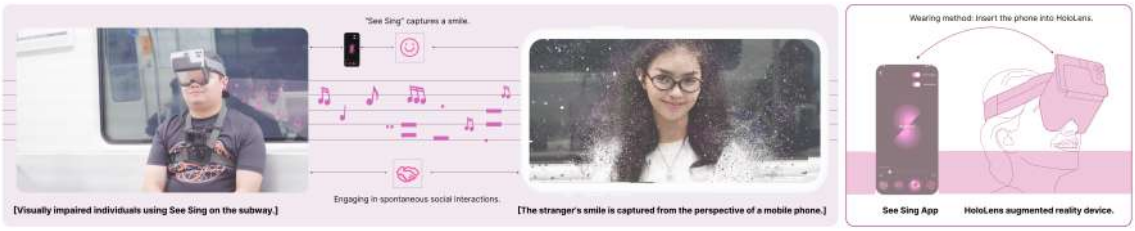


Fig. 1. See Sing——A prototype probe that allows visually impaired individuals to recognize strangers' smiles through the melody of music in spontaneous social situations. When See Sing identifies a smile, it generates a unique melody of the smile personalized by a large model, allowing visually impaired individuals to experience the non-verbal language of smiles in spontaneous social interactions; See Sing requires the use of the HoloKit headset device to enhance reality; See Sing also provides visual effects for individuals with low vision.

Can the visually impaired (VI) "hear" others' smiles? This study employs an ethnographic experiential futures methodology to envision AI-driven melodic auditory augmentation beyond traditional descriptive approaches for spontaneous social interactions involving VI individuals. We propose a mid-fidelity research probe: "See Sing," a smart glasses application that translates facial expressions, gestures, and environmental cues into emotionally contextual soundscapes, akin to real-time cinematic scoring. Our findings indicate that this melodic augmentation not only facilitates VI users' spontaneous interactions with strangers but also enhances their comprehension of emotional nuances in multi-person dialogues. However, the research also unveils ethical concerns, particularly regarding potential misunderstandings arising from AI biases in interpreting social subtleties. This work contributes to the discourse on the intersection of AI, assistive technology, and social interaction in our increasingly tech-mediated world, raising critical questions about the future of sensory augmentation and its societal implications.

CCS Concepts: • **Human-centered computing** → Interaction design theory, concepts and paradigms; **Mixed / augmented reality**; Collaborative and social computing systems and tools; Participatory design.

Additional Key Words and Phrases: Speculative Design, Accessibility, Affective Computing, Visually Impaired, Wearables

ACM Reference Format:

Anonymous Author(s). 2024. Hearing the Smile: Scoring Spontaneous Social Scenes for the Visually Impaired. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (CHI'25).* ACM, New York, NY, USA, 22 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

Manuscript submitted to ACM

1 INTRODUCTION

In [1] the current field of accessibility research within Human-Computer Interaction (HCI), numerous scholars have employed AI technology to achieve significant results in various application scenarios [2, 23, 24, 47, 67, 68] for the visually impaired people(VIs), encompassing everyday life and work. However, research and applications geared towards spontaneous social interactions for VIs are still remarkably scarce.

While some researchers in the field of social interactions for VIs are dedicated to helping them capture subtle social signals [6, 19, 78]—including perceiving the surrounding environment [7, 70, 75], identifying social pauses, and interpreting non-verbal social cues [36, 74] such as facial expressions and body language [14, 57]—other studies [64, 81, 83, 92] explore how VIs can perceive others' gazes [71] or how they can proactively greet weak-tie individuals [92]. For instance, Takayuki Komoda's research assists VIs in interpreting others' smiles and sadness through melodies [39], enhancing their ability to recognize non-verbal signals in social interactions.

However, in initiating spontaneous conversations, especially when encountering strangers, VIs often cannot actively and promptly access potential social signals and struggle to understand nuanced and personalized facial expressions. This results in inconvenience and embarrassment during social interactions due to their inability to accurately perceive the emotions and intentions of others. Consequently, they may lack the initiative and courage to engage socially and may even fear seeking help. As a result of not receiving timely positive emotions conveyed by others, VIs miss opportunities for further communication and interaction.

We envision a future where AI is widely adopted, enabling VIs to move beyond traditional auditory cues to AI-driven descriptive auditory enhancements [81], and ultimately, to more advanced techniques such as soundscapes [50, 55]. These soundscapes would transform subtle emotional cues from others into harmonious, contextually relevant audio experiences. Predictive mapping would further enrich this vision by exploring the characteristics, scenarios, and social dynamics that arise when VIs can perceive their surrounding social environment through sound, facilitating spontaneous conversations. This approach contrasts sharply with traditional research, which tends to focus on structured, planned dialogues.

We focus on the following two research questions:

- Q1: What characteristics, scenarios, and social dynamics emerge when VIs can perceive their surrounding social environment through a melodic augmentation in spontaneous, unscheduled interactions, compared to planned and structured conversations?
- Q2: Will future developments in melodic augmentation enhance the quality of social life for VIs?

We used a hybrid research methodology to answer the research questions, combining Ethnography and Experiential Futures (EXF) [30, 54]. This approach enables us to engage deeply with VIs through formative studies in the early design stages, gaining insights into their needs and expectations. Throughout the design process, we continuously gather feedback to ensure that the final prototype effectively addresses the practical issues and needs of the target users.

In our formative study, we explored various emotional social scenes, encompassing both positive and negative emotions. Interviews indicated that most VIs expressed a desire to initiate spontaneous conversations with strangers. Detecting smiles was identified as particularly effective in enhancing their social life [35, 61]. While smiles can convey a range of underlying emotions and are primarily used to facilitate social openings, extreme negative emotions like sadness or anger, which are less common in everyday interactions, are not typically employed to initiate social engagements.

[1]Since English is not the native language of any author, we employed an AI language model to polish the grammar and enhance the readability of our draft manuscript.

However, these negative emotions can be valuable in fostering empathy, especially in closer relationships, as opposed to interactions with strangers [35, 61]. Based on these findings, we decided to focus on developing a smile detection application as a speculative product to assist VIs in socializing more effectively.

We conducted a low-fidelity online user experiment in which VI participants watched short clips with background music simulating social encounters involving smiles. This was done to gather feedback on their understanding of character expressions and the effects of the accompanying melody. Based on the results of this initial experiment, we focused on developing a smile detection app as a speculative product designed to help VIs socialize more effectively. Subsequent user experiments validated the application's usability and effectiveness across various social scenarios, from spontaneous conversations to daily usage logs, thoroughly assessing the probe's value in social interactions.

We propose a mid-fidelity research tool—a smart glasses application named *"See Sing"*. Through our *"See Sing"* as a future probe, this tool enables visually impaired (VI) users to interpret real-time facial expressions of smiles and contextually emotive soundscapes. The system integrates facial recognition [39, 69, 92], large language models (LLMs) [54], real-time AI music generation [49], and spatial audio [16]. Scenario mapping indicates that this auditory enhancement not only assists VI individuals during spontaneous social interactions with strangers but also deepens their understanding of emotional contexts in multi-person dialogues. This fosters deeper interpersonal connections and contributes to creating a more inclusive and VI-friendly society, potentially progressing toward a utopian scenario.

In this study, we learned that *"See Sing"* offers improvements over traditional descriptive auditory augmenting technologies from a VI's perspective, providing richer storytelling and enhanced effectiveness. However, we also observed widespread mind-blindness caused by AI bias. Our main contributions include:

- C1: Our research has demonstrated significant design potential within the domain of spontaneous social interactions for the visually impaired, highlighting specific scenarios and latent needs.
- C2: Our findings offer several suggestions for future HCI scholars looking to design in this field, indicating areas ripe for further exploration and innovation.

2 RELATED WORKS

2.1 AI-assisted Communication for VI

Many researchers have conducted in-depth studies on the everyday lives and work environments of VIs. In daily life, artificial intelligence technologies assist VIs with various tasks, including navigation [3, 47], travel [1, 8], object recognition [67, 89], and translation [67]. Abigale J. Stangl developed an AI technology that provides VIs with convenient online clothing shopping, improving their shopping experience [80]. In the workplace, Minoli Perera has utilized AI assistants to enhance the productivity and efficiency of blind individuals [68]. In the field of education, AI technologies have also been used to tutor blind children in their daily learning [22]. Beyond these specific domain applications, there are many general-purpose AI assistants [2, 24, 81, 84], such as "Be My Eyes" [5] and Microsoft's "Seeing AI" [68], which enable VIs to perform more complex and diverse tasks. In the social domain, many researchers have alleviated social barriers for VIs through innovations in both virtual and real environments. This includes applications in virtual environments [15], campus environments [48], and weak-tie social interactions [93]. Shaomei Wu used AI to help VIs automatically convert images into textual descriptions online to better understand visual information [90]. The Smart Cane utilizes AI technology to help VIs recognize different faces [32]. Additionally, Sreekar Krishna developed a social assistant to ease the challenges VIs encounter with non-verbal communication in social settings [44]. Separately,

Anam et al. created the Google Glass application, Expression, which aids VIs in social participation by enabling the recognition of facial expressions and body language [4].

In addition to task-specific applications, AI is also used to address challenges in facial expression recognition (FER), a critical area for improving social interactions for VIs. FER aims to identify emotions based on facial movements, but its accuracy is often hampered by biases in AI systems, including racial, gender, and age-based discrimination [18, 91]. Studies have highlighted that AI models tend to misclassify expressions, with smile classifiers being overly sensitive to features associated with youth or assigning negative emotional scores based on race [27, 73]. Moreover, AI's ability to interpret pain through facial expressions is advancing in medical fields, but concerns remain about its reliability and cognitive validity in broader contexts, such as advertising and human resources [21].

Despite these technical challenges, AI holds significant potential to aid VIs in recognizing facial expressions during social interactions, a key component for engaging more meaningfully with others [93]. However, the integration of AI in the social domain has yet to substantially address the more nuanced aspects of interaction, such as sensing environmental cues, understanding social subtleties, or interpreting non-verbal signals like facial expressions and body language. This gap highlights the need for further development and acceptance of AI technologies, which could transform how VIs experience and navigate social environments.

2.2 Non-verbal Communication and VI

2.2.1 Non-verbal Interaction and Social Trust. Many studies have shown that non-verbal information plays a crucial role in social relationships [38]. It is essential for promoting mutual understanding and empathy, establishing trust, and helping individuals gain a sense of belonging, psychological satisfaction, and security [20]. Non-verbal cues provide an elegant means of transmitting, interpreting, and exchanging verbal information. Most non-verbal information is conveyed through communication, with facial expressions being particularly important [38]. Facial expressions are closely related to emotions and are key to conveying emotions and intentions [38, 60]. For example, basic emotions such as "happiness," "sadness," "disgust," "anger," "fear," and "surprise" can be expressed through the shapes of the eyes, mouth, and eyebrows [38, 72]. Non-verbal information can also serve as visual feedback, facilitating empathetic communication. By observing others' facial expressions, people can respond appropriately, such as by returning a smile, thereby promoting empathy and establishing trust.

However, VIs are unable to recognize the facial expressions of their conversation partners, making this form of communication challenging [9, 33, 77]. Research indicates that the absence of non-verbal signals can lead to misunderstandings, anxiety, tension, and even social isolation among VIs, preventing them from establishing social trust [20].

2.2.2 Facial Expressions for the Visual Impaired. In existing research on the transmission and perception of facial expressions [40, 51, 53, 64, 82, 93], many studies have enhanced non-verbal communication through mediated technologies [10, 26]. These studies cover the psychological basis of facial expressions, the biological mechanisms of emotional expression, and facial expression animations in virtual environments [20, 38, 52]. For VIs, the lack of vision makes it difficult to receive non-verbal social signals [9, 33, 77]. To address this, researchers have developed wearable devices and haptic feedback technologies [43, 45] that help VIs better perceive and understand facial expressions.

Some studies have used sound to convert social signals into audio prompts based on facial movements and expressions [37, 40, 51, 85, 93]. For example, the iCare project [45] used camera glasses to describe people's physical characteristics, helping VIs recognize friends. The Expression app [4] for Google Glass assists in social activities by

recognizing facial expressions and body attributes. However, these simple audio prompts often fail to capture the subtleties of emotions.

Tactile substitution technology uses vibration units to translate emotional expressions [53, 58, 64, 82]. For example, eye contact assistance technology [63] uses smart glasses and a haptic wristband to simulate eye contact, and nodding assistance technology mimics nodding actions through smart glasses and a haptic band. Vibro Glove [43] uses vibration motors on fingers to map facial expressions to vibration patterns, showing the potential of haptic feedback in conveying expressions. A haptic chair [10] transmits facial expression information through vibrations on the back. Despite these advancements, challenges remain in portability, learning costs, and conveying subtle emotions.

Some scholars have explored subtle differences in emotional expression further. Takayuki Komoda's auditory interface [39] enables VIs to feel a shared rhythmic melody during interactions, enhancing communication and providing a new emotional experience. An innovative sonification method that assigns colors to instrument sounds [62]provides VIs with a novel means of perceiving visual information through sound. Eye-Phonon [**?**] is a wearable system that adds color and depth to conversations by controlling timbre and pitch.

Our research focuses on the non-verbal signals exhibited by VIs during initial social interactions or encounters, such as smiles and gazes. These signals are crucial for further interactions, and we aim to convey their emotional nuances through musical melodies.

2.2.3 How do VIs interpret film scores to understand a film's emotions? Various interaction technologies have been developed to enhance the movie-watching experience for VIs [17, 31, 55, 59, 65, 79]. Audio descriptions provide detailed narration to help viewers understand character emotions and scene dynamics, especially elaborating on facial expressions and key actions during pauses in dialogue [65, 88]. Yujia Wang et al. implemented automatic descriptions for accessible videos [88**?**]. Music also plays a key role [25, 59] by supplementing the visual information that visually impaired viewers cannot obtain, enhancing the emotional expression of the plot. Stephen James Kroland [46] et al. Designed Automatic Musical Soundscapes of Visual Art for People with Blindness or Low Vision in order to enable blind individuals to better understand complex emotional. Furthermore, haptic technology conveys emotional cues within the movie through touch, such as the proximity of characters or the intensity of scenes, offering visually impaired viewers a richer sensory experience [56**?**]. Researchers like Lakshmie Narayan Viswanathan [87] are using these technologies to make movie actions and emotional atmospheres more vivid.

This study attempts to apply the methods of audio-text emotional descriptions and music-enhanced emotional expressions from movies to the context of improvised, unfamiliar scenarios between VIs and sighted individuals. Through a formative study, we observe the reactions and opinions of VIs and further explore the impact and effectiveness of these interactions through feedback.

2.3 Experiential Futures and Participatory HCI Design Practices in Accessibility Research

2.3.1 Ethnological Experiential Futures. Ethnographic Experiential Futures (EXF) [12, 13] is a design-driven hybrid foresight approach that integrates two research and practice models: Ethnographic Futures Research (EFR) [86] and Experiential Futures (XF) [13]. Its goal is to enhance the accessibility, diversity, and depth of future scenarios. EFR, originating from the formal interview processes pioneered by anthropologist Robert Texler at Stanford University, focuses on descriptively mapping individual or community expectations, fears, and anticipations of the future, thus exploring latent future visions in people's minds [86]. XF employs multisensory, transmedia, and narrative methods to make conceptualized future scenarios visible, tangible, and interactive through future artifacts and immersive

scenarios [13]. EXF combines the strengths of both, not only revealing future scenarios but transforming them into concrete and immersive experiences to foster reflection and change. For instance, Michal Luria and Stuart Candy used the EXF method to explore ethical issues in social agent design through sending letters from the future [54]. Researchers use the EXF framework to interview environmental activists, exploring the relationship between community health and petrochemical industrial facilities [41], and to study the potential futures of supermarkets and their core values [30], thereby creating new future encounters and driving structured reflection and qualitative data analysis.

3 METHODS

To answer our research questions, our study was structured into three stages. In the first stage, we used EXF [12, 13] method to conduct online semi-structured interviews to understand the social challenges faced by VIs and encouraged them to envision future AI-supported social interactions. Our findings revealed that VIs often miss opportunities for proactive communication due to an inability to access non-verbal cues and their desire to initiate spontaneous conversations with strangers. In the second stage, we utilized a Formative Wizard-of-Oz experimental approach, designing a low-fidelity mockup for a "chance encounter" scenario within a film set to simulate the emotional shifts conveyed through real-time music and narrative descriptions. This phase aimed to identify issues faced by VIs in spontaneous social interactions and to validate the importance of auditory cues in such contexts. The third stage involved developing a mid-fidelity smart glasses application, "See Sing," to explore its potential utility in everyday social interactions. During prototype testing, we orchestrated spontaneous social interaction tasks and maintained logs of the participants' usage of *"See Sing"* to evaluate its effectiveness and potential applications. We tested the use of *"See Sing"* under various social scenarios to assess its adaptability for VIs.

This methodology underscores the significance of integrating assistive technologies to enhance the social experiences of visually impaired individuals, providing valuable insights into the dynamics of non-verbal communication and its implications for HCI design.

3.1 Participants

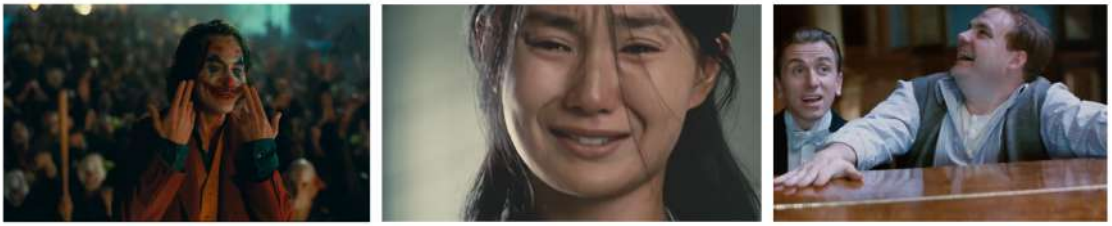
This study recruited six VI participants (2 females, 4 males, Table 1.) including those who are partially sighted and those with low vision, ranging in age from 18 to 40 years old. These participants come from diverse social backgrounds and their occupations include programmers, voice actors, and university students. They are typically active in various social settings and are generally extroverted. Each VI participant is proficient in using smart devices and assistive technologies and has experience with AI products. The selection criteria for participants were based on their familiarity with technology aids and their need for social interaction, aimed at ensuring the broad representativeness and applicability of the research results. Due to the extended duration of our experiment, which was conducted in two phases, we provided each visually impaired participant with a reliable sighted conversation partner to ensure effective results. Fourteen participants (7 pairs) joined our study. All 14 participants received 200 RMB as compensation.

3.2 Study 1: Formative Wizard-of-Oz Experiments to Envision VI Social Interaction in the Future

At the beginning of our experiment, we initiated our formative study by combining ethnography and Experience Futures (EXF), we focused on tailoring future social interaction visions for visually impaired individuals in unfamiliar social settings. The interviews were conducted online via Zoom, each lasting 20 to 30 minutes. We explored the participants' visions of future social interactions involving artificial intelligence, asking them to imagine their expected, unexpected, and imminent future scenarios. We recorded videos of the interview session. We applied Burnard's coding method [11]

ID	Age/Gender	Occupation	Visual Condition
P1	22/M	Student	Ultra low vision
P2	23/M	Student	Low vision
P3	31/M	Voice Actor	Low vision
P4	40/F	Masseur	Blind
P5	29/F	Programmer	Low vision
P6	24/M	Masseur	Blind

Table 1. Demographics of participants.



1. The sarcastic laugh in "Joker"; 2. The sad smile in "Comrades: Almost a Love Story"; 3. The happy smile in "The Legend of 1900".

Fig. 2. Various smiling situations in movie clips used in the study1

to the interview transcripts. Two researchers initially coded separate samples and then compared and discussed the resulting categories. Discussions revolved around issues that might arise in chance encounters, spontaneous social settings, and unfamiliar scenarios, as well as their envisioned futures of social interactions. The purpose of these interviews was to determine the acceptance level of AI-assisted social interactions among visually impaired individuals, explore their deep-seated potential needs, identify pressing issues they face, and translate these issues into specific design directions.

In previous interviews, we noted that visually impaired individuals often refrained from initiating interactions such as casual encounters or conversations in unfamiliar social settings due to their inability to receive non-social signals, such as facial expressions, emotions, and body language. Participants particularly emphasized the importance of being able to hear ambient sounds. In addition to this, we noticed that visually impaired individuals generally have a keen desire to spontaneously interact with strangers and emphasize the crucial role of smile detection in facilitating social interactions. Smiling, as a complex and multifaceted social signal, is primarily used to trigger social interactions, making it easier to initiate conversations in various settings. In contrast, deep negative emotions such as sadness or anger are less common in everyday social interactions and are not typically used to initiate interaction. Although these emotions are extremely important for the development of empathy, especially in established personal relationships, they are not suitable for brief and superficial contact with strangers. Therefore, we adopted the Wizard-of-Oz research method [76] and conducted a low-fidelity experiment featuring movie character encounters. In this experiment, we played approximately 5 minutes of movie clips for each visually impaired participant. These clips were specifically chosen to have no dialogue but included background music illustrating scenes where characters' emotions developed, including three types of smiles, see in Figure 2. Subsequently, we asked the VIs to judge the emotional states of the

Fig. 3. Demonstration of the experimental process; visually impaired individuals complete scenario tasks using *"See Sing"*.

characters in the movie either by listening to the music or through real-time verbal descriptions provided by volunteers via a Zoom meeting.

We designed a mixed-methods experiment aimed at assessing the impact of different assistive tools—text descriptions and music—on the quality of social interactions in various contexts. Using a 2x2 mixed design, we divided six VIs into two groups of three: one group relied on unified text descriptions of the characters' emotional changes narrated by sighted volunteers, while the other group relied on the original movie soundtracks. Each group's experiment was conducted in separate Zoom virtual meeting rooms by a sighted volunteer and a visually impaired participant. After the experiment, the VIs filled out questionnaires to evaluate the effectiveness, comfort, and viewing experience of acquiring non-verbal social information through different methods. Data collection included questionnaire responses and behavioral observations. After the experiment, we conducted about 30 minutes of semi-structured interviews, asking the VIs about the effectiveness of the tools used, and discussing potential designs and applications of these tools in unfamiliar social settings. We hope to evaluate the independent and interactive effects of different variables on the VIs' understanding of non-social signals through this comprehensive approach.

3.3 Study 2: Experiencing a Mid-Fidelity Future – an Probe App *"See Sing"*

In our formative research, we focused on various emotional social scenarios that encompass both positive and negative emotions. Interviews revealed that most visually impaired individuals desire to spontaneously start conversations with strangers. Detecting smiles has proven to be an especially effective way to enhance their social lives [35, 61]. While smiles can convey different underlying emotions, they are primarily used to initiate social interactions. Extreme negative emotions, such as sadness or anger, are less common in everyday life and are usually not employed to initiate social interactions, although these negative emotions are valuable for fostering empathy, especially in closer relationships with strangers [35, 61]. Based on these findings, we decided to focus on developing a speculative product application named "Hearing the Smile – See Sing" to help visually impaired individuals socialize more effectively. This application uses artificial intelligence to generate musical melodies based on detected smiles, with the AI creating melodies that express the emotional nuances of different smiles as accurately as possible.

The *"See Sing"* application is designed following two fundamental principles: Firstly, we aim to enable visually impaired users to perceive the subtle nuances of personalized smiles as much as possible. Secondly, we want visually impaired users to spontaneously hear the melodies of smiles in spontaneous encounters without actively seeking out information.

Fig. 4. Storyboard: spontaneous social interactions in an Elevator.

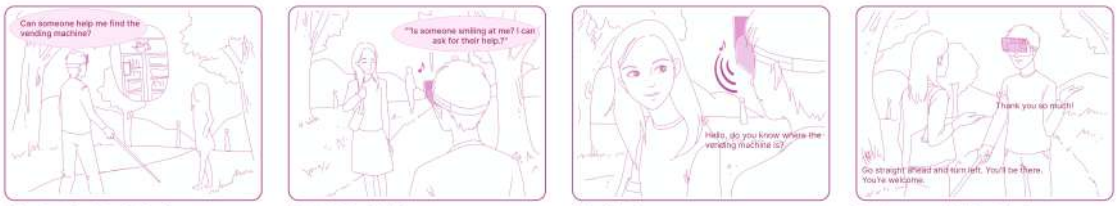


Fig. 5. Storyboard: spontaneous social interactions in street.

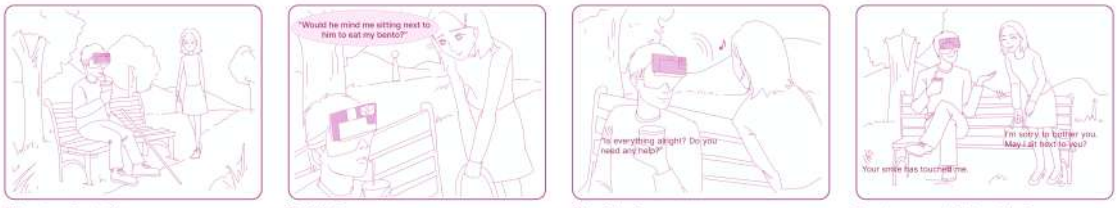


Fig. 6. Storyboard: spontaneous social interactions in park.

Based on the first design principle, we utilize large language models like LLMS to capture smile expressions in unfamiliar scenes through a camera and analyze the smiles' text descriptions with GPT, avoiding the use of any specific prompt words to ensure objectivity and non-interference in the analysis. From these descriptions, we use Suno's API to convert smile expressions into musical melodies, allowing visually impaired users to feel the emotional nuances brought by different smiles. This entire workflow has been integrated into the *"See Sing"* software, with detailed interaction processes available in the design section.

Following the second principle, to ensure that visually impaired users can spontaneously and promptly receive melodies of smiles, we use a device similar to smart glasses to ensure that the camera can capture the moment of a smile in real time, thus not missing any social possibilities. For ease of use, we chose HoloKit—a device that transforms

an iPhone into a VR headset. HoloKit supports spatial audio and requires no charging, working in conjunction with the *"See Sing"* application to serve visually impaired users. Additionally, the application integrates Apple's iOS accessibility features, including the screen reader, to ensure a seamless experience for visually impaired users.

3.3.1 Experiments. In the final research phase, we designed and executed two experiments to assess how the *"See Sing"*, mediated by AI for smile recognition, can facilitate potential social interactions for visually impaired individuals.

We utilized a 2x2 mixed design experiment to evaluate the impact of different assistive tools—text descriptions and music—on the quality of social interactions across various social settings. In the experiment, six VIs were divided into two groups of three: one group used text-based assistive tools, while the other used music-based tools. These participants engaged in several randomly simulated social scenarios, such as buying coffee at a café, commuting in a subway, and strolling in a park (as depicted in Figure 3). In these scenarios, one group of VIs tested using the HoloKit, while the other group used their own familiar AI software for text description recognition.

To add realism to the experiment, we also arranged for six sighted participants to randomly encounter the VIs in these scenarios. These sighted participants were instructed to smile when seeing the VIs, with no restriction on the intensity of the smiles. Through this setup, we aimed to observe the reactions of the VIs during the experiment and evaluate the effectiveness of the *"See Sing"* application.

In the second experiment, we provided six visually impaired individuals with the *"See Sing"* application for them to use freely for a day. We asked the participants to record any new social scenarios facilitated by the application and to keep logs of their usage during the day. We illustrated three storyboards to show the interactions including interacting in an elevator, walking down the street, and strolling in a park as illustrated in Figures 4, 5, and 6.

After the experiment concluded, we conducted one-on-one semi-structured interviews with each participant, with each interview lasting about 30 minutes. During the interviews, we discussed the spontaneous social scenarios encountered during the two experiments and evaluated the effectiveness of the *"See Sing"* application. We recorded videos of the whole experiment sessions. After transcribing these videos, we coded the transcripts using the same method as in Study 1 and identified three themes related to the VI experience when using the *"See Sing"* app.

4 FINDINGS

Based on the results from our Study 1 and 2, as well as the formative Wizard-of-Oz experiments, we have summarized several findings, which can be categorized into four main themes. 1. Potential Applied Social Interaction Scenarios in the future; 2. VI and people's social interaction assisted with *"See Sing"*; 3. Text-based tools vs Sound-based tools for social interaction; 4. Perception and Acceptance of AI-Mediated Discrepancies in Emotional Communication.

4.1 Potential Applied Social Interaction Scenarios in the Future

4.1.1 During their subway commute, VIs may consider using the "See Sing" *app to explore whether anyone is smiling at them.* Many VIs often think of the *"See Sing"* probe during their daily commutes, such as when riding the subway or dining out. They hope to use it to explore whether strangers are smiling at them and to sense the emotional state of those around them, thus gaining more non-verbal social signals at the moment. This perception brings them greater peace of mind during travel and the possibility of social harmony. For example, after our Study2 experiment log ended, through semi-structured interviews, VI2 stated: *"Today, on the way, I met someone who wanted to sit next to me and eat a bento. I invited him to sit down. Before the conversation, I had already felt the other person's friendliness through the friendly smile signal, which reduced my unease. If I could use* "See Sing" *every day when commuting on the subway, I*

would know who smiles at me in the crowded subway, which would greatly relieve my tension and fear of the unknown." VI4 also stated: *"If I could carry this device every day when I go out, I would be more willing to ask others where a good restaurant is, especially those who smile kindly at me, rather than staying at home and ordering takeout, which would increase my possibilities for social interaction."* VI3 mentioned that.*"I am often navigated by maps to unfamiliar streets and need the help of passers-by, but it is difficult to find willing helpers in the crowd."* However, he recorded in his usage log that by identifying smiles, *"I found many friendly strangers willing to help, and initiated positive social interactions, providing him with practical opportunities to ask for directions."* Almost all VIs noted that *"See Sing"*, by recognizing smiles and judging the friendliness of strangers, and providing corresponding social intent musical melodies, greatly assisted them in initiating spontaneous social interactions in different scenarios, offering various forms of help in their current situations.

4.1.2 Using the melodic features feedback from "See Sing" *to remember others.* A VI frequently identified the smile of the same stranger during spontaneous social interactions and encountered this person on different occasions. Since the stranger's way of smiling was essentially the same, the melody heard through the *"See Sing"* device was also similar, allowing the individual to remember this person by the characteristics or tags of the melody. VI6 shared his experience: *"Initially, in the experiment at the cafe, I recognized the person smiling at me through the 'See Sing' device. His music melody was light and gentle, consisting of just one or two chords, which made a deep impression on me. In the afternoon, during our free time in the park, I heard a similar melody. Although I was initially unsure, I realized it was her when she came over to greet me. In the evening, I heard a similar sound again, so I took the initiative to greet her and asked if it was her, which was amazing! I remembered a stranger through the characteristics of the smile melody, and we were very fated."* VI6 also mentioned: *"This way, I can remember many friends and even recognize my clients on the street, remembering different people's basic characteristics through different melodies."* By continuously recognizing the same person's smile and receiving similar melodies through AI, VI6 deepened his memory of the melody, allowing him to proactively greet the person on their third encounter. This became a very unexpected but useful application scenario for "See Sing."

4.1.3 Identifying helpful sighted individuals through"See Sing" *for VIs.* In our scenario experiments, VIs were required to complete various tasks such as finding a coffee shop or shopping using a vending machine. Since VIs cannot rely on smartphone navigation to independently navigate complex road conditions, they often need to ask friendly strangers for help, thus engaging in spontaneous social interactions during their daily travels. However, finding a willing and suitable stranger to help often becomes a challenge they face. In our experimental log, VI5 shared his experience: *"I stood there, heard the melody of a smile, and so I walked towards the direction of the melody. Because he smiled at me first, I knew he would definitely help me, which made me feel very reassured."* Similarly, VI2 successfully completed the task of buying water from a vending machine with the help of a stranger: *"When I was feeling anxious, a melody came from a distance, and I mustered the courage to ask, 'Is there someone over there? Can you help me buy some water?' Usually, I need to wait for someone with a buying need to come close before I can get help. It's not polite and makes me uneasy to shout loudly on the street, but someone smiling at me indicates that they are friendly and kind, which makes me feel very reassured."* Seeking help from strangers on the street is a common situation for many VIs to engage in spontaneous social interactions, although most have experienced being rejected, which can be frustrating and demoralizing. Our *"See Sing"* device can help VIs screen for more friendly and willing strangers in the early stages of spontaneous social interactions, making their social lives more confident, harmonious, and happy.

4.1.4 VIs can use the "See Sing" *device to read the reactions of audience besides the presenter in a multi-person meeting.* In multi-person social settings such as video conferences or offline meetings, VIs often face gaps in social interactions. Through our experiments, it was found that the *"See Sing"* device can help VIs receive non-verbal signals in multi-person interactions, enabling them to better understand the context of conversations. For example, they could perceive audience reactions, even if the audience did not make any sound. VI1 shared his experience during the experiment, noting that when he received a call from a friend, he could hear his friend's smile through *"See Sing"*, even though there was no sound on the phone. *"Previously, when friends called me, it was often just for company, so they wouldn't speak. I used to guess what they were doing and whether my words made them unhappy. Sometimes they would have other friends on the call, but I wouldn't know. Now, with this device, I can easily sense that two people are smiling at me because I hear two different melodies."* Similarly, in offline interactions, VI4 encountered a similar situation, *"While I was listening to a staff member explain precautions, a waiter brought coffee and smiled at me. I interrupted the presenter to ask if the waiter was smiling at me. This made me feel capable of constantly monitoring everyone's reactions around me, although this can sometimes be distracting."* With *"See Sing"*, VIs are able to capture immediate feedback from different people in multi-person social scenes, even taking initiative in conversations, significantly enhancing their social abilities and sense of participation.

4.1.5 VIs use"See Sing" *to gauge others' emotions.* Our experiments also demonstrated that many VIs could use *"See Sing"* to gauge the emotions of others in spontaneoussocial settings. In interviews, most VIs mentioned that the degree of variation in the melodies enhanced their judgment. VI3 shared his experience: *"While I was talking to a stranger in the park, their music changed from a soothing melody at the beginning to a very intense pop music rhythm. Later, he even laughed out loud, and I could feel that his laugh was very different from the previous melody; I could even detect his fake laugh because it didn't match the earlier melody, and of course, his laughter gave him away. 'See Sing' enhanced my judgment."* Additionally, VIs could also sense the emotional feedback of people around them in multi-person social settings. For example, VI3 stated: *"When I was speaking, I noticed that they were all smiling at me, and I felt that they must be very interested in what I was saying. Previously in such situations, I would worry about not speaking well and not understanding why they weren't speaking, but now I can know their feedback through the melodies of smiles."* The experimental results also showed that the different melodies of smiles identified by AI helped VIs understand others' emotions. VI6 heard two different smile melodies while walking in the park, *"I chose the more urgent, faster one; I guessed he might know me, otherwise why would he smile so brightly at me?"* By recognizing complex smiles through AI and displaying the real-time personalized melody differences of different individuals, VIs can more finely perceive others' emotions, thereby gaining richer social signals and possibilities in spontaneous social interactions.

4.2 VI and people's social interaction assisted with *"See Sing"*

4.2.1 VIs using "See Sing" *can raise their eyebrows at strangers before starting a conversation.* In our experiments, we observed that VIs were able to utilize *"See Sing"* to capture non-verbal signals, such as smiles, before the start of spontaneous social interactions and thus initiate subtle social cues towards strangers, including eyebrow raises and body movements. For example, VI3 was resting on a park bench when a stranger wanted to sit next to her. Despite the stranger standing far away and the surrounding bird calls drowning out any sounds, VI3 could capture the melody of the smile through *"See Sing"*. She stated, *"I heard her smile melody and waved at her, which made me feel very liberated."* Another VI participant didn't make an obvious social gesture but mentioned his reaction during an interview: *"I heard the melody, though I wasn't sure if I heard it right, so I just raised my eyebrows to show that I had received the signal. This*

way, it wasn't awkward, and it was also polite." Although VIs responded to the melodies of smiles with varying degrees of body language and facial expressions, these actions demonstrate that *"See Sing"* can empower them to initiate subtle interactions before spontaneous social engagements, thereby establishing weak social connections.

4.2.2 Using "See Sing" *has filled the once quiet world of VIs with sounds of friendliness.* Using *"See Sing"* has enabled VIs to experience unexpected joy in their lives, increasing their reception of kind sounds, subtle encouragements, and support from society during their daily outings. VI3 shared his feelings: *"This product has brought me unexpected happiness because I can sense more people's smiles. It informs me of others' laughter in a new way, allowing me to proactively feel it without needing someone else to convey it."* VI1 also expressed similar emotions: *"This product makes me feel more people's smiles, bringing unexpected joy."* VI2 added: *"The smiles conveyed by the product have given me unexpected happiness in everyday life."* Moreover, VI1 noted: *"This product gives me more courage when interacting with strangers."* The majority of VIs have experienced unexpected happiness through the smiles conveyed in this way. While this sense of happiness may vary with different situations, the melodies of smiles have enhanced their ability to engage in spontaneous social interactions with society.

4.2.3 Using "See Sing" *in multi-person conversations can lead to awkward interactions for VIs.* During our experiments, we also discovered that using *"See Sing"* could not only facilitate positive and harmonious interactions but also lead to potential conflicts, which are themselves a part of the interaction process. For example, VI5, after identifying a smile in a group chat and moving closer to the source of the melody, inadvertently made the other person feel uncomfortable and awkward. We observed and asked VI5 about his perspective on this incident. He explained, *"I thought he wanted to talk to me, so I moved closer, but it might have just been a polite smile."* This misunderstanding made him hesitant in subsequent experiments, cautious about initiating interactions, *"I apologized, but it wasn't intentional."* Although there was no overt conflict, this incident still created an unpleasant atmosphere. We documented this conflict, indicating that using *"See Sing"* is a double-edged sword: while it allows VIs to engage in spontaneous social scenarios based on the melodies of smiles, these interactions can sometimes be unpleasant.

4.2.4 Using "See Sing" *has helped VIs make more friends.* To our surprise, the logs from users of *"See Sing"* revealed that VIs had added many new friends on social media, far exceeding their usual frequency of making new friends. For instance, VI1, while strolling in the park, enthusiastically shared our product's features and his personal experiences with strangers. During the conversation, the strangers became very interested in his life and added him on social media. *"A group of people added me on social media, and I felt very special. I was confident and happy to find that initiating conversations with strangers isn't scary at all, and everyone's smiles were friendly."* Similarly, VI2 received social media invitations from two strangers: *"I heard their smiles' melody from across the street, and I asked them if it was a red light, they said it was green. After that, we started chatting."* In various social settings, VIs, hearing the melody of smiles through *"See Sing"*, were able to turn spontaneous social interactions into friendships. This positive feedback demonstrates the significant potential of this technology in enhancing the social lives of VIs.

4.3 Text-based tools vs Sound-based tools for social interaction

4.3.1 Text-based tools can be untimely and awkward, while sound-based tools are natural and less awkward. In our initial formative studies, most VIs expressed reluctance to use text-based tools for identifying others' facial expressions during social interactions, considering it both awkward and unnatural. They often had to guess others' reactions. VI3 shared his experience: *"When I argue with my family, the house becomes very quiet, and I can't discern their emotions and state.*

I'm too embarrassed to use text-based tools for identification; it feels very unnatural and awkward. I hope AI can solve more of these non-verbal signals, so I can understand more information timely." In our Formative Wizard-of-Oz experiments, VI5 commented: *"It was too awkward when I tried to engage with the movie content, and the staff attempted to describe it but interrupted the background music, and the plot had already changed; I felt like my movie experience was dropped frames."* VI6 stated: *"For such delicate expressions of emotions, we generally don't describe them, just quietly listen to the music and feel it. It's the most primal, natural experience, and we can actually feel many emotions through the music."* Most VIs prefer describing delicate emotions and moods through music, and using text-based tools in spontaneous social scenarios not only poses timeliness issues but may also make others feel odd, thus causing unnecessary conflicts. VI4 added: *"We would never pick up our phones to scan what's happening around us; it would be embarrassing. Even if we did, we might miss timely reactions."* VIs hope to recognize social signals while avoiding awkwardness and unnaturalness, finding that experiencing non-verbal signals through musical melodies is very suitable and provides a refined and natural social experience.

4.3.2 Using text-based tools can lead to auditory information overload. In our formative "Wizard-of-Oz" experiments, most VIs reported missing parts of movie scenes when using text-based tools, as this added to the auditory information burden, leading to information overload and preventing them from clearly hearing the dialogue. VI2 shared his perspective: *"If I had to choose, I would definitely prefer using music to describe subtle social signals, because music is inherently an additional mode of information. Using text descriptions now actually disrupts my understanding of the plot."* VI5 also commented: *"Using text tools in actual conversations is even more unreliable, as it can interrupt my communication and even disrupt my thought process just as social interactions are beginning."* Therefore, most VIs do not wish to rely on AI's text-based tools for recognizing non-verbal social signals, as these not only interrupt their interactions but also make it difficult to distinguish between primary and secondary information, leading to information overload.

4.3.3 Sound-based tools allow for simultaneous perception of feedback from other audience members in the meeting. In multi-person communications, most VIs struggle to simultaneously notice everyone's emotions and feedback. They often can only focus on the presenter's speech, which frequently leads them to feel isolated and unable to deeply integrate into the group communication environment. VI5 shared their feelings: *"Due to visual impairments, I hate noisy environments and most of the time can only focus on one person speaking. At work, I often overlook my colleagues' feelings, especially those with normal vision, which leads to poor relationships. They often have to accommodate me, and it's hard for me to establish deep communications with them, which makes me very distressed."* VI6 expressed hope for AI technology: *"I hope AI can tell me everyone's emotional state without interfering with the presenter's speech. I think musical melodies are great, like watching a movie with both dialogues and background music, which lets me know when the story moves to the next scene."* VI3 also mentioned the delicate depiction through music: *"Refined music allows every character in a movie to be detailed, even insignificant passersby, as long as there's their shot and background music, I can feel subtle emotional fluctuations. There's no such technology in real life, making me feel like I'm living in a movie."* VI1 shared observations on Korean dramas: *"When protagonists in Korean dramas meet, there is often background music, which feels very dreamy. But when I meet strangers on the street, I only feel nervous and scared because I can't sense whether they are like the protagonists in the movies. If there were sound-based tools, they could help me better integrate into these spontaneous scenes."* In early formative studies, VIs imagined sound-based tools that could score most of the non-verbal signals in social processes like movie music and delicately capture each person's emotional changes without missing any crucial information.

4.4 Perception and Acceptance of AI-Mediated Discrepancies in Emotional Communication

During the experiment, five VIs generally accepted the discrepancies between the AI-generated smile melodies and actual smiles. Two sighted individuals also acknowledged these biases but did not attribute them to AI. VI4 commented, *"Even if there are discrepancies between the melody of a recognized smile and the actual image or emoji, it doesn't impact our communication with society."* VI1 challenged, *"Even if you can see the real world, aren't there errors? So why can't I accept the biases introduced by AI? I think these biases are individual, not AI-induced."* VI2 added, *"I can accept the discrepancies between the pictures and AI music melodies, as there are similarities."* VI6 noted, *"As long as the discrepancy is not completely contradictory, I can accept it; at least it's not deceptive."* VI6 believed that communication could still be effective despite inaccuracies: *"Even if there are errors, as long as both parties are willing to communicate, there is no problem. We might also misunderstand things ourselves; I experience misunderstandings in my conversations with friends too."* VI3 suggested that *"these biases are primarily due to the absence of a shared textual context between visually impaired and sighted individuals."* As VIs receive more information in the future, errors will be corrected, discrepancies will narrow, and 'mind-blindness' will diminish. In discussions between sighted individuals and VIs, both groups acknowledged that everyone has biases in understanding things; AI does not provide completely contrary musical melodies. During the smile description process, both parties expressed excitement and anticipation of each other's responses, leading to vibrant discussions. Despite differences in how smiles were described, there was a keen interest in understanding each other's perspectives.

5 DISCUSSION

5.1 See Sing used in various spontaneous social interaction

In our research, VIs demonstrated a variety of surprising functions and applications using *"See Sing"* in spontaneous social settings. For example, *"See Sing"* allows visually impaired individuals to pay attention to non-verbal signals during daily outings, thereby regaining a sense of lost happiness, helping them make new friends in unfamiliar environments, identify willing and friendly helpers, and even judge others' emotional feedback based on personalized melodies or remember others by melody features. Additionally, *"See Sing"* is used to read reactions from audience members other than the main speaker in multi-person meetings. According to Norman's psychological theory [66], "users often don't know what they need," many of the needs and application scenarios for spontaneous social interactions were not fully perceived by VIs before we developed "See Sing." However, through our research, these application scenarios and functions have facilitated deeper connections between visually impaired individuals and people, environments, life, and work in society to varying degrees, and even opened up the possibility for visually impaired individuals to experience their surroundings in an entirely new way. This approach is not limited to "hearing the smile" in spontaneous social situations but also includes experiencing the world through soundscapes, such as feeling the sounds of landscapes, exploring melodies between different cities, enjoying the melodic textures of different leaves, and listening to the unique rhythms at sunrise each day. These findings reveal a vast design space and numerous potential application scenarios waiting for future HCI researchers to further explore and develop.

5.2 Comparison between different assistive tools for VIs

As a sound-based tool, we are more suited for spontaneous social scenarios compared to most text-based tools. Our tool allows visually impaired individuals to naturally receive more non-verbal social signals in spontaneous settings, which are often difficult to convey through text. In experiments, we compared *"See Sing"* with AI text description

software [4, 31, 81] commonly used by visually impaired individuals, and the results showed that most do not prefer to proactively use text description tools in spontaneous social interactions. Even if visually impaired individuals cannot engage in spontaneous social interactions visually, they still anticipate encountering the melody of smiles in social settings. This has also been validated by visually impaired individuals in our research findings, marking a significant advantage for spontaneous social scenarios.

Secondly, while many HCI scholars study tactile sensory interaction design to compensate for the reception of non-verbal social signals [43, 53, 64, 82], this type of reception often lacks the granularity needed for spontaneous, personalized, complex, and diverse emotional expressions. Most tactile interactions are relatively limited. "See Sing," through its AI model, can capture different strangers' smiles and generate different musical melodies based on the intensity and style of each person's smile, allowing VIs to remember strangers' smiles through personalized melodies, much like remembering a friend's hobbies and habits.

Lastly, in our preliminary research, a few researchers have used musical melodies to help visually impaired individuals receive non-verbal signals in regular conversations, such as Takayuki Komoda's study [39] which assists them in interpreting smiles and sadness through melodies. Although this enhances their ability to recognize non-verbal signals in social interactions, listening to melodies for a long time during conversations might be distracting. Therefore, we focus this functionality at the start of spontaneous scenarios, identifying strangers' smiles, and concentrate more on using AI to generate personalized character melodies to provide visually impaired individuals with distinct memory tags. This not only improves the recognition of non-verbal signals in social interactions but also emphasizes the importance as a channel and bridge for communication and integration with society.

In summary, we have shifted the trajectory of our research on the application of non-verbal social signals from focusing on conversations between visually impaired individuals and their friends to how they might engage in spontaneous social interactions with the wider society, strangers, and new acquaintances. *"See Sing"* serves as the initial connecting bridge, contributing to the body of knowledge for HCI researchers.

5.3 Limitations

AI bias may increase misunderstandings between sighted and visually impaired individuals, widening the gap between them. AI biases often arise from the selection of training datasets that typically favor the majority, leading to representative bias [93]. However, the acceptance and trust in AI-assisted facial expression recognition, especially among different user groups, remains unclear. In our formative study, we observed that most VIs were willing to accept a certain level of discrepancy between the music melodies generated by AI to recognize smiley emojis and actual human smiles. We refer to this phenomenon as "mind's eye blindness." When asked about their willingness to accept this "blindness," most visually impaired respondents expressed a positive attitude, acknowledging that errors are inevitable in the development of technology and can be optimized over time.

In contrast, sighted individuals held differing opinions regarding this acceptance by VIs. As noted by Engelmann et al. [21], concerns were raised about the cognitive validity of AI inferences, leading to doubts about its reliability in interpreting facial expressions. Some sighted participants found it difficult to accept that visually impaired individuals could tolerate these discrepancies, fearing that such tolerance could result in the miscommunication of subtle details in future interactions. They were particularly concerned that this might widen the social gap between sighted and visually impaired people.This difference in attitudes highlights a critical issue: the acceptance of technological biases and errors is uneven across user groups. Such discrepancies in acceptance may contribute to misunderstandings and increased biases in social interactions between sighted and visually impaired individuals. To address this issue, it is important

to incorporate educational initiatives and awareness campaigns into the design and implementation of AI systems, ensuring that all users better understand the limitations and potential biases of the technology.

In this study, we delve into the complex dynamics of entanglement[34] that arise when artificial intelligence serves as a technological intermediary in social interactions for visually impaired individuals. While these dynamics have facilitated further integration of visually impaired people with society to some extent, they have also introduced "psychic blindness," where visually impaired individuals may overlook or misinterpret certain social cues while relying on AI for social interaction. As designers in Human-Computer Interaction (HCI), our responsibility is to thoughtfully intervene in these complex entanglements with a more human-centered approach to design.

AI bias is distributed among VIs, society, and manufacturers. The differing levels of bias acceptance between visually impaired (VI) individuals and sighted people pose an intriguing question: Who should be empowered to address and mitigate bias? Should the authority to debias AI systems be granted to VIs who directly experience the technology's limitations, or should it remain with sighted individuals who may place a different emphasis on the accuracy of non-verbal communication? This question underscores the complexity of designing inclusive AI systems, highlighting the need for a balanced approach that considers the perspectives and needs of all user groups.

During software development, developers can reduce biases by adjusting parameters. Additionally, users can refine the AI systems they use through continuous training. Political correctness at a societal level also plays a role in mitigating bias. Our research indicates that visually impaired (VI) individuals are willing to learn about the potential drawbacks of technological advancements because, in their assessment, using assistive tools like *"See Sing"* enhances their social interactions. They can even debug these tools themselves to better suit their needs. However, *"See Sing"* still occasionally misinterprets awkward smiles as genuine, benevolent smiles. In multi-person interactions, sighted individuals may explain to VIs that an awkward smile depicted in an emoticon is not meant to be benevolent, thereby adjusting *"See Sing"*'s recognition outcomes. Therefore, if political correctness is pursued at a societal level, it is essential to consider whether this aligns with VIs' expectations. Whether they prefer to receive specialized treatment or be treated like sighted individuals remains an issue that needs further exploration.

5.4 Design Implications

Using a magnifying glass to focus on the bridge between VI and society. In current research, most AI applications targeting the visually impaired concentrate on their daily living and work, appearing superficially similar to HCI applications aimed at the general public. However, beyond studying visually impaired or other special groups in isolation, perhaps we can shift our focus to their relationships and connections with society. For example, how can visually impaired individuals toast and dine with sighted individuals without awkwardness? How can they participate in sports together to deepen mutual understanding and communication? Or how can they dance together to showcase their unique movements and dance styles? These scenarios contain rich design spaces waiting for more HCI scholars to develop and explore in the future.

Our research conclusions and findings suggest that probes like "See Sing," as transitional design [28] prototypes, can help visually impaired individuals transition from a supported group to equal participants in a society where everyone is recognized and harmonious [29, 42]. This empowerment of individuals is also an inclusion and advancement of society as a whole. The rapid development of technology allows us to greatly magnify the needs in the link between visually impaired individuals and society, opening up further avenues for exploration and impact.

Hearing the world becomes possible. In our final design of "See Sing," we specifically chose smile recognition as a trigger for potential friendly interactions in spontaneous social settings. However, looking to the future, VIs will be able to customize which non-verbal social cues to use as triggers, such as nodding, the duration of a gaze, or frowning. These non-verbal cues help visually impaired individuals make more accurate social judgments.

Additionally, these non-verbal signals can be adjusted according to different application scenarios: for example, in potentially dangerous environments, a prolonged gaze might be more critical; in family activities, rich body language might be more emphasized. By adjusting and applying these emotional signals, visually impaired individuals can better navigate and interact in various environments.

Moreover, expanding the recognition of non-verbal social signals to a broader range of applications, such as through soundscape technology, VIs in the future could "hear" different landscapes and cultures, like the sound of plants growing, the subtle changes of sunrise and sunset, or the tranquil atmosphere in temples. This is not just recognition of non-verbal signals but a whole new sensory experience, opening a vast space for design exploration, allowing visually impaired individuals to feel and understand the world in completely new ways.

6 CONCLUSIONS

We propose a mid-fidelity research tool—a smart glasses application called *"See Sing"*. Through our *"See Sing"* as a future probe, this tool allows visually impaired (VI) users to interpret real-time facial expressions of smiles and contextually emotive soundscapes. The system integrates facial recognition, large language models (LLMs), real-time AI music generation, and spatial audio. Scenario mapping shows that this auditory enhancement aids VIs in spontaneous social interactions with strangers and enriches their understanding of emotional contexts in multi-person dialogues. This fosters deeper interpersonal connections and contributes to a more inclusive and VI-friendly society, potentially progressing towards a utopian scenario.

In this study, we learned that *"See Sing"* outperforms descriptive auditory augmenting technologies from a VI's perspective, offering more storytelling and effectiveness. However, we also observed the widespread mind-blindness caused by AI bias. Our research has demonstrated significant design potential within the domain of spontaneous social interactions for the visually impaired, highlighting specific scenarios and latent needs. Additionally, our findings provide actionable insights for future HCI scholars aiming to innovate in this field, pinpointing areas ripe for further exploration and innovation.

REFERENCES

[1] Ali Abdolrahmani, Maya Howes Gupta, Mei-Lian Vader, Ravi Kuber, and Stacy Branham. 2021. Towards More Transactional Voice Assistants: Investigating the Potential for a Multimodal Voice-Activated Indoor Navigation Assistant for Blind and Sighted Travelers. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 495, 16 pages. https://doi.org/10.1145/3411764.3445638

[2] Ali Abdolrahmani, Ravi Kuber, and Stacy M. Branham. 2018. "Siri Talks at You": An Empirical Investigation of Voice-Activated Personal Assistant (VAPA) Usage by Individuals Who Are Blind. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility* (Galway, Ireland) *(ASSETS '18)*. Association for Computing Machinery, New York, NY, USA, 249–258. https://doi.org/10.1145/3234695.3236344

[3] Dragan Ahmetovic, Cole Gleason, Chengxiong Ruan, Kris Kitani, Hironobu Takagi, and Chieko Asakawa. 2016. NavCog: a navigational cognitive assistant for the blind. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Florence, Italy) *(MobileHCI '16)*. Association for Computing Machinery, New York, NY, USA, 90–99. https://doi.org/10.1145/2935334.2935361

[4] ASM Iftekhar Anam, Shahinur Alam, and Mohammed Yeasin. 2014. Expression: A dyadic conversation aid using Google Glass for people who are blind or visually impaired. In *6th International Conference on Mobile Computing, Applications and Services*. IEEE, 57–64.

[5] Mauro Avila, Katrin Wolf, Anke Brock, and Niels Henze. 2016. Remote assistance for blind users in daily life: A survey about be my eyes. In *Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments*. 1–2.

[6] Erin Brady, Meredith Ringel Morris, Yu Zhong, Samuel White, and Jeffrey P. Bigham. 2013. Visual challenges in the everyday lives of blind people. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 2117–2126. https://doi.org/10.1145/2470654.2481291

[7] Stacy M. Branham and Shaun K. Kane. 2015. Collaborative Accessibility: How Blind and Sighted Companions Co-Create Accessible Home Spaces. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 2373–2382. https://doi.org/10.1145/2702123.2702511

[8] Andrius Budrionis, Darius Plikynas, Povilas Daniušis, and Audrius Indrulionis. 2022. Smartphone-based computer vision travelling aids for blind and visually impaired individuals: A systematic review. *Assistive Technology* 34, 2 (2022), 178–194.

[9] HP Buimer, M Bittner, T Kostelijk, TM van der Geest, A Nemri, RJA van Wezel, and Y Zhao. 2018. Conveying facial expressions to blind and visually impaired persons through a wearable vibrotactile device. *Plos one* 13, 3 (2018), e0194737–e0194737.

[10] Hendrik Buimer, Thea Van Der Geest, Abdellatif Nemri, Renske Schellens, Richard Van Wezel, and Yan Zhao. 2017. Making Facial Expressions of Emotions Accessible for Visually Impaired Persons. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, Baltimore Maryland USA, 331–332. https://doi.org/10.1145/3132525.3134823

[11] Philip Burnard. 1991. A method of analysing interview transcripts in qualitative research. *Nurse education today* 11, 6 (1991), 461–466.

[12] Stuart Candy. 2010. The futures of everyday life: Politics and the design of experiential scenarios. *University of* (2010).

[13] Stuart Candy and Kelly Kornet. 2019. Turning Foresight Inside Out: An Introduction to Ethnographic Experiential Futures. *Journal of Futures Studies* 23, 3 (2019).

[14] Shonal Chaudhry and Rohitash Chandra. 2015. Design of a mobile face recognition system for visually impaired persons. *arXiv preprint arXiv:1502.00756* (2015).

[15] 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* (New York, NY, USA) *(ASSETS '23)*. Association for Computing Machinery, New York, NY, USA, Article 38, 14 pages. https://doi.org/10.1145/3597638.3608386

[16] Gustavo Corrêa De Almeida, Vinicius Costa de Souza, Luiz Gonzaga Da Silveira Júnior, and Maurício Roberto Veronez. 2024. Spatial Audio in Virtual Reality: A systematic review. In *Proceedings of the 25th Symposium on Virtual and Augmented Reality* (Rio Grande, Brazil) *(SVR '23)*. Association for Computing Machinery, New York, NY, USA, 264–268. https://doi.org/10.1145/3625008.3625042

[17] Khang Dang, Hamdi Korreshi, Yasir Iqbal, and Sooyeon Lee. 2023. Opportunities for Accessible Virtual Reality Design for Immersive Musical Performances for Blind and Low-Vision People. In *Proceedings of the 2023 ACM Symposium on Spatial User Interaction* (Sydney, NSW, Australia) *(SUI '23)*. Association for Computing Machinery, New York, NY, USA, Article 33, 21 pages. https://doi.org/10.1145/3607822.3614540

[18] Iris Dominguez-Catena, Daniel Paternain, and Mikel Galar. 2024. Metrics for Dataset Demographic Bias: A Case Study on Facial Expression Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2024).

[19] Laura E Dreer, Timothy R Elliott, Donald C Fletcher, and Marsha Swanson. 2005. Social problem-solving abilities and psychological adjustment of persons in low vision rehabilitation. *Rehabilitation Psychology* 50, 3 (2005), 232.

[20] Paul Dumouchel. 2005. Trust as an action. *European Journal of Sociology/Archives européennes de sociologie* 46, 3 (2005), 417–428.

[21] Severin Engelmann, Chiara Ullstein, Orestis Papakyriakopoulos, and Jens Grossklags. 2022. What people think AI should infer from faces. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 128–141.

[22] Shalini Garg and Shipra Sharma. 2020. Impact of artificial intelligence in special need education to promote inclusive pedagogy. *International Journal of Information and Education Technology* 10, 7 (2020), 523–527.

[23] Ricardo Gonzalez, Jazmin Collins, Shiri Azenkot, and Cynthia Bennett. 2024. Investigating Use Cases of AI-Powered Scene Description Applications for Blind and Low Vision People. *arXiv preprint arXiv:2403.15604* (2024).

[24] Ricardo E Gonzalez Penuela, Jazmin Collins, Cynthia Bennett, and Shiri Azenkot. 2024. Investigating Use Cases of AI-Powered Scene Description Applications for Blind and Low Vision People. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 901, 21 pages. https://doi.org/10.1145/3613904.3642211

[25] Achim Hättich and Martina Schweizer. 2020. I hear what you see: Effects of audio description used in a cinema on immersion and enjoyment in blind and visually impaired people. *British Journal of Visual Impairment* 38, 3 (2020), 284–298.

[26] Liwen He, Yifan Li, Mingming Fan, Liang He, and Yuhang Zhao. 2023. A Multi-modal Toolkit to Support DIY Assistive Technology Creation for Blind and Low Vision People. In *Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) *(UIST '23 Adjunct)*. Association for Computing Machinery, New York, NY, USA, Article 3, 3 pages. https://doi.org/10.1145/3586182.3616646

[27] Ben Hutchinson, Emily Denton, Margaret Mitchell, and Timnit Gebru. 2019. Detecting bias with generative counterfactual face attribute augmentation. In *Proceedings of the Fairness, Accountability, Transparency and Ethics in Computer Vision Workshop*.

[28] Terry Irwin, Cameron Tonkinwise, and Gideon Kossoff. 2022. Transition design: An educational framework for advancing the study and design of sustainable transitions. *Cuadernos del Centro de Estudios en Diseño y Comunicación. Ensayos* 105 (2022), 31–72.

[29] Russell Jacoby. 2005. *Picture imperfect: Utopian thought for an anti-utopian age*. Columbia University Press.

[30] Tom Jenkins, Laurens Boer, Juliane Brigitta Busboom, and Ivar Østby Simonsen. 2020. The Future Supermarket: A Case Study of Ethnographic Experiential Futures. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society* (Tallinn, Estonia) *(NordiCHI '20)*. Association for Computing Machinery, New York, NY, USA, Article 77, 13 pages. https://doi.org/10.1145/3419249.3420130

[31] Lucy Jiang, Crescentia Jung, Mahika Phutane, Abigale Stangl, and Shiri Azenkot. 2024. "It's Kind of Context Dependent": Understanding Blind and Low Vision People's Video Accessibility Preferences Across Viewing Scenarios. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 897, 20 pages. https://doi.org/10.1145/3613904.3642238

[32] Yongsik Jin, Jonghong Kim, Bumhwi Kim, Rammohan Mallipeddi, and Minho Lee. 2015. Smart Cane: Face Recognition System for Blind. In *Proceedings of the 3rd International Conference on Human-Agent Interaction* (Daegu, Kyungpook, Republic of Korea) *(HAI '15)*. Association for Computing Machinery, New York, NY, USA, 145–148. https://doi.org/10.1145/2814940.2814952

[33] Divya Jindal-Snape. 2004. Generalization and maintenance of social skills of children with visual impairments: Self-evaluation and the role of feedback. *Journal of Visual Impairment & Blindness* 98, 8 (2004), 470–483.

[34] Deborah G Johnson and Mario Verdicchio. 2024. The sociotechnical entanglement of AI and values. *AI & SOCIETY* (2024), 1–10.

[35] Lucy Johnston, Lynden Miles, and C Neil Macrae. 2010. Why are you smiling at me? Social functions of enjoyment and non-enjoyment smiles. *British Journal of Social Psychology* 49, 1 (2010), 107–127.

[36] Katherine Mary Jones, Ute Leonards, and Oussama Metatla. 2024. "I Don't Really Get Involved In That Way": Investigating Blind and Visually Impaired Individuals' Experiences of Joint Attention with Sighted People. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 52, 16 pages. https://doi.org/10.1145/3613904.3642940

[37] Mohammad Kianpisheh, Franklin Mingzhe Li, and Khai N. Truong. 2019. Face Recognition Assistant for People with Visual Impairments. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 3 (Sept. 2019), 1–24. https://doi.org/10.1145/3351248

[38] Mark L Knapp, Judith A Hall, and Terrence G Horgan. 1978. *Nonverbal communication in human interaction*. Vol. 1. Holt, Rinehart and Winston New York.

[39] Takayuki Komoda, Hisham Elser Bilal Salih, Tadashi Ebihara, Naoto Wakatsuki, and Keiichi Zempo. 2024. Auditory Interface for Empathetic Synchronization of Facial Expressions between People with Visual Impairment and the Interlocutors. In *Proceedings of the Augmented Humans International Conference 2024* (Melbourne, VIC, Australia) *(AHs '24)*. Association for Computing Machinery, New York, NY, USA, 138–147. https://doi.org/10.1145/3652920.3652937

[40] Takayuki Komoda, Hisham Elser Bilal Salih, Tadashi Ebihara, Naoto Wakatsuki, and Keiichi Zempo. 2024. Auditory Interface for Empathetic Synchronization of Facial Expressions between People with Visual Impairment and the Interlocutors. In *Proceedings of the Augmented Humans International Conference 2024*. ACM, Melbourne VIC Australia, 138–147. https://doi.org/10.1145/3652920.3652937

[41] Kelly Kornet. 2015. Causing An Effect: Activists, Uncertainty & Images of the Future. (2015).

[42] Gideon Kossoff. 2015. Holism and the reconstitution of everyday life: A framework for transition to a sustainable society. *Design Philosophy Papers* 13, 1 (2015), 25–38.

[43] Sreekar Krishna, Shantanu Bala, Troy McDaniel, Stephen McGuire, and Sethuraman Panchanathan. 2010. VibroGlove: An Assistive Technology Aid for Conveying Facial Expressions. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*. ACM, Atlanta Georgia USA, 3637–3642. https://doi.org/10.1145/1753846.1754031

[44] Sreekar Krishna, Dirk Colbry, John Black, Vineeth Balasubramanian, and Sethuraman Panchanathan. 2008. A systematic requirements analysis and development of an assistive device to enhance the social interaction of people who are blind or visually impaired. In *Workshop on Computer Vision Applications for the Visually Impaired*.

[45] Sreekar Krishna, Greg Little, John Black, and Sethuraman Panchanathan. 2005. iCARE Interaction Assistant: A Wearable Face Recognition System for Individuals with Visual Impairments. In *Proceedings of the 7th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, Baltimore MD USA, 216–217. https://doi.org/10.1145/1090785.1090837

[46] Stephen James Krol, Maria Teresa Llano, Matthew Butler, and Cagatay Goncu. 2024. Design Considerations for Automatic Musical Soundscapes of Visual Art for People with Blindness or Low Vision. *arXiv preprint arXiv:2405.14188* (2024).

[47] Bineeth Kuriakose, Raju Shrestha, and Frode Eika Sandnes. 2023. Exploring the User Experience of an AI-based Smartphone Navigation Assistant for People with Visual Impairments. In *Proceedings of the 15th Biannual Conference of the Italian SIGCHI Chapter* (Torino, Italy) *(CHItaly '23)*. Association for Computing Machinery, New York, NY, USA, Article 17, 8 pages. https://doi.org/10.1145/3605390.3605421

[48] Amanda Lannan. 2019. A Virtual Assistant on Campus for Blind and Low Vision Students. *Journal of Special Education Apprenticeship* 8, 2 (2019), n2.

[49] Adam Hollmén Larsen and Jichen Zhu. 2024. Ideary: Facilitating Electronic Music Creation with Generative AI. In *Companion Publication of the 2024 ACM Designing Interactive Systems Conference* (IT University of Copenhagen, Denmark) *(DIS '24 Companion)*. Association for Computing Machinery, New York, NY, USA, 275–278. https://doi.org/10.1145/3656156.3663731

[50] Mark Lawton, Stuart Cunningham, and Ian Convery. 2020. Nature soundscapes: an audio augmented reality experience. In *Proceedings of the 15th International Audio Mostly Conference* (Graz, Austria) *(AM '20)*. Association for Computing Machinery, New York, NY, USA, 85–92. https://doi.org/10.1145/3411109.3411142

[51] Kyungjun Lee, Daisuke Sato, Saki Asakawa, Chieko Asakawa, and Hernisa Kacorri. 2021. Accessing Passersby Proxemic Signals through a Head-Worn Camera: Opportunities and Limitations for the Blind. In *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, Virtual Event USA, 1–15. https://doi.org/10.1145/3441852.3471232

[52] Anping Liu and Hongjie Yue. 2024. Facial Expression Recognition Based on CNN-LSTM. In *Proceedings of the 2023 7th International Conference on Electronic Information Technology and Computer Engineering* (Xiamen, China) *(EITCE '23)*. Association for Computing Machinery, New York, NY, USA, 486–491. https://doi.org/10.1145/3650400.3650480

[53] Leon Lu, Jin Kang, Chase Crispin, and Audrey Girouard. 2023. Playing with Feeling: Exploring Vibrotactile Feedback and Aesthetic Experiences for Developing Haptic Wearables for Blind and Low Vision Music Learning. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility* (New York, NY, USA) *(ASSETS '23)*. Association for Computing Machinery, New York, NY, USA, Article 14, 16 pages. https://doi.org/10.1145/3597638.3608397

[54] Michal Luria and Stuart Candy. 2022. Letters from the Future: Exploring Ethical Dilemmas in the Design of Social Agents. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 419, 13 pages. https://doi.org/10.1145/3491102.3517536

[55] Keenan R. May, Brianna J. Tomlinson, Xiaomeng Ma, Phillip Roberts, and Bruce N. Walker. 2020. Spotlights and Soundscapes: On the Design of Mixed Reality Auditory Environments for Persons with Visual Impairment. *ACM Trans. Access. Comput.* 13, 2, Article 8 (apr 2020), 47 pages. https://doi.org/10.1145/3378576

[56] Antonella Mazzoni and Nick Bryan-Kinns. 2016. Mood glove: A haptic wearable prototype system to enhance mood music in film. *Entertainment Computing* 17 (2016), 9–17.

[57] Troy McDaniel, Diep Tran, Samjhana Devkota, Kaitlyn DiLorenzo, Bijan Fakhri, and Sethuraman Panchanathan. 2018. Tactile Facial Expressions and Associated Emotions toward Accessible Social Interactions for Individuals Who Are Blind. In *Proceedings of the 2018 Workshop on Multimedia for Accessible Human Computer Interface* (Seoul, Republic of Korea) *(MAHCI'18)*. Association for Computing Machinery, New York, NY, USA, 25–32. https://doi.org/10.1145/3264856.3264860

[58] Troy L. McDaniel, Daniel Villanueva, Sreekar Krishna, Dirk Colbry, and Sethuraman Panchanathan. 2010. Heartbeats: a methodology to convey interpersonal distance through touch. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI EA '10)*. Association for Computing Machinery, New York, NY, USA, 3985–3990. https://doi.org/10.1145/1753846.1754090

[59] Christina E Mediastika and Anugrah S Sudarsono. 2020. Sound matters while enjoying movies; a soundscape study of visually impaired people. In *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, Vol. 261. Institute of Noise Control Engineering, 3599–3608.

[60] Albert Mehrabian. 1981. Silent messages: Implicit communication of emotions and attitudes.

[61] Daniel Messinger and Alan Fogel. 2007. The interactive development of social smiling. *Advances in child development and behaviour* 35 (2007), 328–366.

[62] Haruna Miyakawa, Hisham Elser Bilal Salih, Tadashi Ebihara, Naoto Wakatsuki, and Keiichi Zempo. 2022. Eye-Phonon: Wearable Sonification System based on Smartphone that Colors and Deepens the Daily Conversations for Person with Visual Impairment. In *Adjunct Publication of the 24th International Conference on Human-Computer Interaction with Mobile Devices and Services* (Vancouver, BC, Canada) *(MobileHCI '22)*. Association for Computing Machinery, New York, NY, USA, Article 17, 6 pages. https://doi.org/10.1145/3528575.3551449

[63] Lauren Murray, Philip Hands, Ross Goucher, and Juan Ye. 2016. Capturing Social Cues with Imaging Glasses. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*. ACM, Heidelberg Germany, 968–972. https://doi.org/10.1145/2968219.2968260

[64] Isabel Neto, Yuhan Hu, Filipa Correia, Filipa Rocha, Guy Hoffman, Hugo Nicolau, and Ana Paiva. 2024. Conveying Emotions through Shape-changing to Children with and without Visual Impairment. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 49, 16 pages. https://doi.org/10.1145/3613904.3642525

[65] Zheng Ning, Brianna L Wimer, Kaiwen Jiang, Keyi Chen, Jerrick Ban, Yapeng Tian, Yuhang Zhao, and Toby Jia-Jun Li. 2024. SPICA: Interactive Video Content Exploration through Augmented Audio Descriptions for Blind or Low-Vision Viewers. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 902, 18 pages. https://doi.org/10.1145/3613904.3642632

[66] Donald A Norman and Andrew Ortony. 2003. Designers and users: Two perspectives on emotion and design. In *Symposium on foundations of interaction design*. Interaction Design Institute, 1–13.

[67] Nadia Oukrich, Bougary Tamega, and Naziha Laaz. 2023. Matia Application: An AI Multi-Lingual Assistant For Visually Impaired And Blind People. In *Proceedings of the 6th International Conference on Networking, Intelligent Systems & Security* (Larache, Morocco) *(NISS '23)*. Association for Computing Machinery, New York, NY, USA, Article 6, 7 pages. https://doi.org/10.1145/3607720.3607727

[68] Minoli Perera. 2024. Enhancing Productivity Applications for People who are Blind using AI Assistants. In *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems (CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 432, 6 pages. https://doi.org/10.1145/3613905.3638180

[69] Leandro Persona, Fernando Meloni, and Alessandra Alaniz Macedo. 2023. An accurate real-time method to detect the smile facial expression. In *Proceedings of the 29th Brazilian Symposium on Multimedia and the Web* (Ribeirão Preto, Brazil) *(WebMedia '23)*. Association for Computing Machinery, New York, NY, USA, 46–55. https://doi.org/10.1145/3617023.3617031

[70] Lope Ben Porquis, Sara Finocchietti, Giorgio Zini, Giulia Cappagli, Monica Gori, and Gabriel Baud-Bovy. 2017. ABBI: A wearable device for improving spatial cognition in visually-impaired children. In *2017 IEEE biomedical circuits and systems conference (BioCAS)*. IEEE, 1–4.

[71] Shi Qiu, Siti Aisyah Anas, Hirotaka Osawa, Matthias Rauterberg, and Jun Hu. 2016. E-Gaze Glasses: Simulating Natural Gazes for Blind People. In *Proceedings of the TEI '16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction*. ACM, Eindhoven Netherlands, 563–569.

https://doi.org/10.1145/2839462.2856518

[72] Shi Qiu, Jun Hu, and Matthias Rauterberg. 2015. Nonverbal signals for face-to-face communication between the blind and the sighted. In *Proceedings of International Conference on Enabling Access for Persons with Visual Impairment*. 157–165.

[73] Lauren Rhue. 2018. Racial influence on automated perceptions of emotions. *Available at SSRN 3281765* (2018).

[74] Sharon Sacks, Linda Kekelis, and Robert Gaylord-Ross. 1992. *The development of social skills by blind and visually impaired students: Exploratory studies and strategies*. American Foundation for the Blind.

[75] Sharon Zell Sacks, Karen E Wolffe, and Deborah Tierney. 1998. Lifestyles of students with visual impairments: Preliminary studies of social networks. *Exceptional Children* 64, 4 (1998), 463–478.

[76] Daniel Salber and Joëlle Coutaz. 1993. Applying the wizard of oz technique to the study of multimodal systems. In *Human-Computer Interaction: Third International Conference, EWHCI'93 Moscow, Russia, August 3–7, 1993 Selected Papers 3*. Springer, 219–230.

[77] Lei Shi, Brianna J. Tomlinson, John Tang, Edward Cutrell, Daniel McDuff, Gina Venolia, Paul Johns, and Kael Rowan. 2019. Accessible Video Calling: Enabling Nonvisual Perception of Visual Conversation Cues. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 131 (nov 2019), 22 pages. https://doi.org/10.1145/3359233

[78] Pranali Uttam Shinde and Aqueasha Martin-Hammond. 2024. Designing to Support Blind and Visually Impaired Older Adults in Managing the Invisible Labor of Social Participation: Opportunities and Challenges. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 50, 14 pages. https://doi.org/10.1145/3613904.3642203

[79] Joel Snyder. 2005. Audio description: The visual made verbal. In *International congress series*, Vol. 1282. Elsevier, 935–939.

[80] Abigale J. Stangl, Esha Kothari, Suyog D. Jain, Tom Yeh, Kristen Grauman, and Danna Gurari. 2018. BrowseWithMe: An Online Clothes Shopping Assistant for People with Visual Impairments. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility* (Galway, Ireland) *(ASSETS '18)*. Association for Computing Machinery, New York, NY, USA, 107–118. https://doi.org/10.1145/3234695.3236337

[81] Vincent Stragier, Omar Seddati, and Thierry Dutoit. 2023. Developing an Interactive Agent for Blind and Visually Impaired People. In *Proceedings of the 2023 ACM International Conference on Interactive Media Experiences* (Nantes, France) *(IMX '23)*. Association for Computing Machinery, New York, NY, USA, 248–253. https://doi.org/10.1145/3573381.3596471

[82] Zhe Sun, Robin Ananda, and Xinyi Fu. 2022. EmoSparkle: Tangible Prototype to Convey Visual Expressions for Visually Impaired Individuals in Real-time Conversations. In *Proceedings of the Tenth International Symposium of Chinese CHI*. ACM, Guangzhou, China and Online China, 38–49. https://doi.org/10.1145/3565698.3565768

[83] Zhe Sun, Robin Ananda, and Xinyi Fu. 2024. EmoSparkle: Tangible Prototype to Convey Visual Expressions for Visually Impaired Individuals in Real-time Conversations. In *Proceedings of the Tenth International Symposium of Chinese CHI* (Guangzhou, China and Online, China) *(Chinese CHI '22)*. Association for Computing Machinery, New York, NY, USA, 38–49. https://doi.org/10.1145/3565698.3565768

[84] Nourhan Tahoun, Anwar Awad, and Talal Bonny. 2020. Smart Assistant for Blind and Visually Impaired People. In *Proceedings of the 3rd International Conference on Advances in Artificial Intelligence* (Istanbul, Turkey) *(ICAAI '19)*. Association for Computing Machinery, New York, NY, USA, 227–231. https://doi.org/10.1145/3369114.3369139

[85] Md. Iftekhar Tanveer, A.S.M. Iftekhar Anam, A.K.M Mahbubur Rahman, Sreya Ghosh, and Mohammed Yeasin. 2012. FEPS: A Sensory Substitution System for the Blind to Perceive Facial Expressions. In *Proceedings of the 14th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, Boulder Colorado USA, 207–208. https://doi.org/10.1145/2384916.2384956

[86] Robert B Textor. 1995. The ethnographic futures research method: An application to Thailand. *Futures* 27, 4 (1995), 461–471.

[87] Lakshmie Narayan Viswanathan. 2011. *Enhancing Movie Comprehension For Individuals Who Are Visually Impaired Or Blind Through Haptics*. Arizona State University.

[88] Yujia Wang, Wei Liang, Haikun Huang, Yongqi Zhang, Dingzeyu Li, and Lap-Fai Yu. 2021. Toward Automatic Audio Description Generation for Accessible Videos. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 277, 12 pages. https://doi.org/10.1145/3411764.3445347

[89] Linda Yilin Wen, Cecily Morrison, Martin Grayson, Rita Faia Marques, Daniela Massiceti, Camilla Longden, and Edward Cutrell. 2024. Find My Things: Personalized Accessibility through Teachable AI for People who are Blind or Low Vision. In *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems (CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 403, 6 pages. https://doi.org/10.1145/3613905.3648641

[90] Shaomei Wu, Jeffrey Wieland, Omid Farivar, and Julie Schiller. 2017. Automatic Alt-text: Computer-generated Image Descriptions for Blind Users on a Social Network Service. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 1180–1192. https://doi.org/10.1145/2998181.2998364

[91] Tian Xu, Jennifer White, Sinan Kalkan, and Hatice Gunes. 2020. Investigating bias and fairness in facial expression recognition. In *Computer Vision–ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part VI 16*. Springer, 506–523.

[92] Yuhang Zhao, Shaomei Wu, Lindsay Reynolds, and Shiri Azenkot. 2018. A Face Recognition Application for People with Visual Impairments: Understanding Use Beyond the Lab. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–14. https://doi.org/10.1145/3173574.3173789

[93] Yuhang Zhao, Shaomei Wu, Lindsay Reynolds, and Shiri Azenkot. 2018. A face recognition application for people with visual impairments: Understanding use beyond the lab. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–14.