Volume 11
May 2023
ISSN: 2330-4216

2023

The Journal on Technology and Persons with Disabilities

Journal Track Proceedings, Online 2023

CSUN | CENTER ON DISABILITIES

Journal on Technology and Persons with Disabilities

ISSN 2330-4216

LIBRARY OF CONGRESS * U.S. ISSN CENTER
ISSN Publisher Liaison Section
Library of Congress
101 Independence Avenue SE
Washington, DC 20540-4284
(202) 707-6452 (voice); (202) 707-6333 (fax)
issn@loc.gov (email); www.loc.gov/issn (web page)

The 38th Annual CSUN Assistive Technology Conference and the 11th Volume of the Journal

The Center on Disabilities at California State University, Northridge is proud to welcome you to the eleventh issue of the *Journal on Technology and Persons with Disabilities*. For over ten years, the journal has been an important forum for academics and researchers to showcase their work and highlight advances in the field. The Journal is comprised of published proceedings from the Annual CSUN Assistive Technology Conference, representing sessions from the Journal Track.

Over the last three decades, the conference has become the most significant global platform for meeting and exchanging ideas. The CSUN AT Conference is committed to bringing the conference to everyone, supporting research and projects, and the value they contribute to knowledge, new innovations and best practices to promote inclusion for all.

We are proud to serve the science and research community by providing the open access publishing opportunity to all presented works. All works included in the *Journal on Technology and Persons with Disabilities* were rigorously reviewed and accepted by both the Journal Track Review Committee and our esteemed Journal Panel of Chairs.

A special thank you to all the authors, the Journal Track Review Committee, Journal Panel of Chairs, the CSUN Center on Disabilities team, and the editorial staff for their professional support. As always, we are grateful for and appreciate the many participants and partners who have contributed to the success of the CSUN Assistive Technology Conference and the Journal.

Regards,

Klaus Miesenberger
Chair, Journal Track
Scientific Editor, *Journal on Technology and Persons with Disabilities*

Julia Santiago
Interim Managing Director, Center on Disabilities
Managing Editor, *Journal on Technology and Persons with Disabilities*

Journal on Technology and Persons with Disabilities Contributors

Center on Disabilities
California State University, Northridge
18111 Nordhoff Street
Northridge, California 91330-8340

Table of Contents

THE JOURNAL ON TECHNOLOGY AND PERSONS WITH DISABILITIES

Triangulating Stakeholder Perceptions of a Higher Education Disability Support Center

Alan M. Safer, Lesley S.J. Farmer
California State University Long Beach
alan.safer@csulb.edu, lesley.farmer@csulb.edu

Abstract

Most studies of higher education disability services focus on one group of stakeholders, but few triangulate the perceptions of multiple stakeholders. To increase disability service center (DSC) effectiveness and increase constituent satisfaction at a large urban public university, surveys were administered using the software Qualtrics to describe the shared and different perceptions of all those involved, including the following: students, DSC staff, and faculty. A key perception related to improved effectiveness that all parties agreed upon was: the encouragement of systematic and respectful knowledge-based communication and accountable action. This factor requires the DSC to conduct greater outreach. On their part, students and faculty agreed on the need to be more mindful and deliberate in accessing needed training and resources, and then carrying out responsible action. The survey findings further suggest the need for more staff and faculty accountability and more student input on DSC services.

Keywords

Data analytics, disability service centers, faculty, students, staff, perceptions.

Introduction

Since most studies about perceptions of higher education disability services focus on one group of stakeholders (Claxton 2017; Scott & Ashmore 2020), this research will focus on the perceptions of multiple stakeholder groups. This research was performed using a number of surveys at a major US public university. The survey data specifically related to the perceptions of students, staff and faculty relative to the operations of the disability service center (DSC) as they experience them directly. The ultimate aim of the surveys was to compare group similarities and differences in the perceptions of DSC services.

Literature Review

A review of the literature of DSC studies revealed that perceptions can vary significantly according to stakeholder affiliation. For instance, students' perceptions of DSCs showed a number of barriers: confusion and lack of knowledge about eligibility and available services (Mamboleo et al. 2020), concern about policy and procedure compliance (Scott 2021), lack of DSC support in transitioning to the workplace (Shpigalman et al. 2021), and a sense of stigma about requesting DSC support (Corby, Taggart and Cousins 2020). Furthermore, students gained academic learning strategies and self-advocacy skills from DSCs (Scott 2021). Shpigalman et al. (2021) also found that students considered DSCs as good resources to help them navigate institutional bureaucratic processes as well as support their personal development and emotional well-being.

Faculty had mixed feelings about DSCs. Similar to students, faculty reported having insufficient knowledge about DSC services and accommodations; it appeared that campus disability service providers did not convey information effectively, and that faculty tended to get their information elsewhere (Harbour and Greenberg 2017; Thurston et al. 2017). Supporting the importance of those findings, McCarron (2020) asserted that knowledge was the most important factor leading to faculty actions supporting college students with disabilities; such faculty knew what to do and how to do it, had realistic expectations about those efforts, and were willing to make those efforts. Echoing McCarron, Young, Schaefer and Lesley (2019) found that teacher training and collaboration with the DSC were key to effective accommodations.

In assessing their own effectiveness, disability service center staff noted the importance of educating faculty (Galkin 2017). While documentation was deemed not as important, DSC staff appreciated timely reviews and asserted that inconsistencies across reviewers and staff burnout negatively affected the process (Miller et al. 2019). To address that issue, Wooderson, Cuskelly and Meyer (2014) found that combining workplace environment improvement with staff development was effective. DSC participants in Prohn, Kelley and Westling's 2019 study remarked on the importance of the campus environment, although many parties had little knowledge or experience working with students who had disabilities. The DSC staff felt that they facilitated those students' social inclusion through personal relationships and helping those students expand their social networks.

Across the board at the macro level, studies reveal that most stakeholder entities mention the importance of knowledge and skills, and the need for training in those areas. They also mention the lack of resources and time to implement accommodations effectively. The social experiences and development of students with disabilities is also considered very important, and studies mention the need to provide a more accepting and supportive inclusive campus climate. In that respect, effective communication and collaboration with all stakeholders, including campus service providers, are needed.

Methods

The basic survey questions raised were:

- How do college students who receive disability services perceive the DSC?
- How do faculty perceive the quality of DSC services they enable?
- How do DSC staff perceive the value of the DSC services they implement?

To answer these research questions, the investigators gathered information about the perceptions of the stakeholders -- the DSC staff and the students and faculty they directly serve -- and supplemented their opinions with other available data about the DSC and campus. To this end, the investigators developed and emailed Qualtrics surveys targeted to all of the members of each group at the investigators' university. Each data set was analyzed separately using descriptive statistics, various multivariate analyses including chi-square analysis, correlation analysis and other techniques; finally, text analytics and inductive content analysis were used for open-ended responses. Afterwards, the three data sets were compared for possible commonalities and differences.

Results

Demographic information about each stakeholder group was collected as follows:

- Of the 689 active students receiving DSC services, 139 responded to the survey in full. Questions addressed student participation and satisfaction with services. This included what additional programs and services they would like.
- Of the 2216 faculty members who had contact with DSC, 399 responded to the survey in full. Questions addressed faculty satisfaction with services, degree of confidence in providing accommodations, their level of knowledge about university legal obligations in providing accommodations, the impact of distance learning on

providing accommodations, sources of information about disabilities and associated accommodations, and additional training they want.

- Of the 22 staff members, 16 responded to the survey in full. Questions addressed staff job satisfaction, what functions and skills were performed most often, their technological skill level, their training, their level of stress, and areas for improvement.

Survey Findings

The major research findings were:

- Student satisfaction: one-time use of DSC and ongoing use of DSC correlated with greater satisfaction of DSC. Students wanted additional programs and mentoring to complement individual staff appointments, especially the longer they received services. Students wanted more DSC staff communication and guidance and wanted those staff to collaborate more with faculty and campus service providers. They also wanted longer operating hours and more drop in options.
- Faculty with more experience with disabilities in their life were more comfortable implementing accommodations and had higher expectations of the DSC. Part-time faculty received less information and training. All faculty were most challenged in dealing with students who had behavioral difficulties or sensory impairments. Faculty also had difficulties modifying assignments and changing due dates. In terms of training, faculty were most interested in learning about assistive technology, dealing with behavioral difficulties, and knowing policies and procedures related to disabilities. They also wanted to establish a communications committee as a number of faculty had difficulties working with the DSC.

- Staff enjoyed working with people and did not enjoy administrative tasks. To better help clientele more, staff wanted more uninterrupted time, more collaboration with the campus community, and clearer expectations about workload and processes. The staff also wanted more training and resources. Many of the staff felt a lot of stress in their duties dealing with students and faculty. Staff also emphasized students' and faculty's own responsibilities in terms of being knowledgeable about disabilities and proactively requesting and providing accommodations.

All stakeholder groups thought that DSC could help students succeed academically. Indeed, seventy percent of students who responded stated that they would highly recommend DSC to a friend. The survey delved into the details of the experiences and attitudes of the respondents, especially in the open-ended questions. As more nuanced findings emerged, each group revealed unique perspectives. Furthermore, multivariate analyses identified interesting relationships.

- The most common accommodation that faculty provided was extended test time. Lecture recording, note-taking, and excused absences were also frequently mentioned.
- Staff spent more of the time working with students on explaining procedures, identifying disabilities and relevant accommodations.

Discussion

A key factor for improved DSC effectiveness that was agreed upon by all parties was: systematic and respectful knowledge-based communication and accountable action. This factor requires DSC staff to conduct greater outreach with training and easily available resources--including going directly to college and department meetings. Similarly, students and faculty need to be more mindful and deliberate in accessing training and available resources, and then carrying out their responsibilities. Some specific recommendations based on the survey findings included providing longer operating hours and drop-in options, providing more training, establishing a communications committee, and developing a mentor program.

Recommendations

The DSC director and her supervisor can explain to the campus leadership about the data-based need for providing accommodations and supporting students with disabilities (as evidenced in this study). Strong encouragement from those leaders is needed to support their own faculty in this effort—which involves active participation in training such as universal design for learning.

Another feasible step is to establish a communications committee with student, faculty, DSC and other campus service provider representatives who can facilitate ongoing communication with constituents. That committee can also review existing documentation and processes to identify gaps in information and bottlenecks in processes. Here, varied perspectives can facilitate needed clarification that can result in less confusion and more compliance.

There is much work to do, and it needs to be done systematically over time with the willing participation of all stakeholders and the support of campus leadership. DSC planning needs to determine priorities and a feasible timeline to be implemented by sufficient staff allocation of effort and needed resource, including time and compensation.

Follow-up Actions Taken

The study was presented to the DSC director and her supervisor. As a starter, the director requested the campus governance to establish a communications committee, and that is to happen soon. To address staff and faculty's challenges in dealing with behavior problems, DSC provided a workshop for them. The counselling center can help in this too. More systematically in fall 2022 DSC implemented Faculty Welcome Meetings for all colleges, which overviewed DSC services and procedures, including faculty and student responsibilities. It is recommended that these meetings continue to occur throughout the year along with having data collected on the amount of people attending along with feedback from faculty to continuously improve communications.

As another means of information and support, the DSC website included additional information for students, such as alternative testing processes and tips for succeeding in virtual courses (which constituted an accommodation for some students). The updated website provides more resources for faculty, such as tips for classroom and exam accommodations, university design for learning and accessible instructional materials, and working with hearing impaired students. The website also has advice for parents of students with disabilities about to enter college. As another technology advance, DSC is working with the campus's Academic Technology Services to implement university for learning across the campus; one achievement is ReadSpeaker, which allows the campus learning management system to provide text-to-speech for students within the courses and during lockdown exams. We are highly recommending that website analytics be stored to keep track of the number of people utilizing the website and which parts of the website are being used (from semester to semester to determine trends to find if the website is being utilized much over time) and with feedback from users on what was helpful and what can be improved. Finally, the director has agreed to the investigator's suggestion from the survey results to implementing a mentoring program for the students at DSC. Other schools have successfully implemented such mentoring programs of students at DSC and the investigators have set up a minimal cost process with mentors utilizing the college of education at the university. The usefulness of each of these recommendations can only be determined if data is collected at each step, along with feedback from those involved, for continuous improvement.

Conclusion

The study demonstrates how systematically gathering and analyzing data, including triangulating perspectives, not only gives a fuller picture of disability services, but it also reveals assumptions and underlying needs and desires that can unite constituencies and facilitate efforts to continuously improve services to students with disabilities. Because campus administrators do not provide accommodations directly to students, they were not surveyed in this study, However, the evidence-based data does provide a strong case to give to campus leaders, calling on them to facilitate concerted efforts to make constituents aware and knowledgeable – and to provide the training, human and material resources, and accountability measures to make sure everyone will responsibly do their part to make sure DSC students succeed.

Works Cited

Corby, Deidre, Laurence Taggart, and Wendy Cousins. “The Lived Experience of People with Intellectual Disabilities in Post-Secondary or Higher Education.” *Journal of Intellectual Disabilities*, vol. 24, no. 3, 2020, pp. 339-357.

Galkin, Daphne L. “Beyond Accommodations: Disability Service Professionals in Higher Education as Campus Change Agents”. *Dissertation Abstracts International Section A: Humanities and Social Sciences*, vol. 76, issue 9–A(E), 2016.

Harbour, Wendy S., and Daniel Greenberg. “Campus Climate and Students with Disabilities.” *NCCSD Research Brief*, vol. 1, no. 2, 2017, vtechworks.lib.vt.edu/bitstream/handle/10919/86931/StudentsDisabilities.pdf?sequence=1. Accessed 29 September 2022.

Mamboleo, George, Shengli Dong, Sean Anderson, and Adrionia Molder. “Accommodation Experience: Challenges and Facilitators of Requesting and Implementing Accommodations among College Students with Disabilities.” *Journal of Vocational Rehabilitation*, vol. 53, no. 1, 2020, pp. 43-54.

McCarron, Elizabeth C. “Postsecondary Faculty and Willingness to Provide Academic Accommodations for Students with Learning Disabilities.” *Journal of Postsecondary Education and Disability*, vol. 34, no. 4, 2020, pp. 347-367.

Miller, Sydney, Ryan Zayac, Amber Paulk, and Stacy Lee. “Disability Accommodation Requests: Prevalence and Preference of Review Processes at Postsecondary Institutions in the United States.” *Journal of Postsecondary Education and Disability*, vol. 32, no. 3, 2019, pp. 217-226.

Prohn, Seb M., Kelly R. Kelley, and David L. Westling. “Supports' Perspectives on the Social Experiences of College Students with Intellectual Disability.” *Inclusion*, vol. 7, no. 2, 2019, pp. 111-124.

Scott, Sally. *The 2020 Biennial AHEAD Survey: Reporting on Disability Resource Professionals in Higher Education*. Huntersville, NC: Association on Higher Education and Disability, 2021.

Shpigelman, Carmit-Noa., Sagit Mor, Dalia Sachs, and Naomi Schreuer. “Supporting the Development of Students with Disabilities in Higher Education: Access, Stigma, Identity, and Power.” *Studies in Higher Education*, 2021, pp. 1–16.

Thurston, Linda, Cindy Shuman, B. Jan Middendorf, and Cassandra Johnson. “Postsecondary STEM Education for Students with Disabilities: Lessons Learned from a Decade of NSF Funding.” *Journal of Postsecondary Education and Disability*, vol. 30, no. 1, 2017, pp. 49-60.

Wooderson, John, Monica Cuskelly, and Kim Meyer. “A Systematic Review of Interventions for Improving the Work Performance of Direct Support Staff.” *Research and Practice in Intellectual and Developmental Disabilities*, vol. 1, no. 2, 2014, pp. 160-173.

Young, Davey, Matthew Schaefer, and Jamie Lesley. "Accommodating Students with Disabilities Studying English as a Foreign Language (Practice Brief)." *Journal of Postsecondary Education and Disability*, vol. 32, no. 3, 2019, pp. 311-319.

Task Aware Browsing: AI Assisted Web Access for Blind Users

David Cane
Independent Researcher
davidcane@alum.mit.edu

Abstract

A new approach for blind users to access web pages is presented. It uses machine learning and artificial intelligence (AI) to overcome problems with conventional browsers and screen readers, particularly on an unfamiliar web site. A set of typed commands is provided to allow a task-oriented approach to completing tasks on a web application. Rather than having to look for a search box, the user need only type "search," or its equivalent hotkey. Rather than tabbing through a page of clutter, only the results of the search are provided. It uses AI to convert the spatial layout of active controls and the subsequent results, to a linear dialog model. The new approach is called Task Aware Browsing, TAB for short.

Work performed to date indicates that machine learning techniques will allow this approach to work on a significant number of web sites. User testing suggests that it provides an appreciable improvement in the speed with which blind users can accomplish desired tasks compared to existing screen readers.

This design also allows for speech assistants to work on any site that TAB can work on, without the need of specific scripts for that site.

Keywords

Web applications, visually impaired, command line interface, artificial intelligence, machine learning.

Introduction

Problem Statement

The PC era began with a command line interface, not a graphical one. Screen readers made the PC generally accessible for blind users who could touch type.

The advent of graphical user interfaces (GUI), while generally a boon to sighted users, presented a variety of significant challenges for blind users. Commands were now found in two dimensions with a pointing device, and output appeared in a variety of places on the screen. The usage pattern for blind users was to use shortcut keys whenever possible, and tab and arrow keys to locate menu items for which shortcut keys were not available. To a certain extent, application developers followed consistent patterns of menu layout and shortcut keys, which reduced the learning time of users for new applications.

Web based applications were appreciably more difficult. There is no longer a consistent design approach for different applications, even those in the same domain. Screen readers provide a variety of mechanisms for users to find the element they are looking for. Also present are mechanisms to search for keywords on the page such as looking for “search” to find the text entry element that invokes a search on the site. That does not work, of course, on a site that chooses to label that element “find”. The display from the result of a command is often filled with clutter. Users need to repeatedly tab though elements, seeking the requested information.

There are accessibility guidelines (W3C), primarily intended to make sure that a screen reader can properly describe the user activated elements on the screen, so they can be found. It was determined from a focus group held with six employees, at Perkins School for the Blind, that there was a significant difference between *accessibility* and *usability*. A site might meet all of the accessibility guidelines of W3C which guarantee that all elements can be found. But it might still be difficult to perform the operations that were desired because it was hard to find what was needed, even though it was possible. They also pointed out that it was not unusual to spend 30-60 minutes learning the layout of a new web site before much could be accomplished with it.

The TAB Concept

Consider how we all bought airplane tickets before the advent of the Internet. We picked up the phone and had a dialog with an agent about where and when we wanted to go. The agent read back the choices and we made a selection. The process was identical for a blind person, or a deaf one with a TTY: linear conversation with a series of commands and responses. Task Aware Browsing (TAB) uses AI to enable that process on a variety of web sites.

We categorize web sites into a collection of vertical domains. Each domain contains web sites that provide the same kind of capabilities; e.g. shopping, travel, banking, Customer Relationship Management, research. For each domain, a set of commands is designed to enable the operations that a user would need to complete the tasks needed in that domain. TAB takes each command and performs that operation on the web site. The following shows the style of a sample TAB dialog for shopping. Greater details will be provided in the Discussion section.

Table 1. TAB dialog style.

User types	**Speech Response**
search eggs	Description of each of the first 5 results from searching for eggs
add 2 3rd	Confirms two items of third product described added to cart.
cart	Describes all items in the cart and total cost
checkout	Confirms checkout in progress, gives total cost

Discussion

Related Work

Implementation of TAB requires effective interface design and semantic web analysis. Most current literature on user interface design focuses on GUI design, ignoring command line design issues. Alonso et al. discuss issues of dual interface design, for both sighted and blind users and provides some insights into design concepts for command line overlays for GUI applications.

Pages generally have some structure to allow sighted users to find what they are looking for and previous work in automated extraction of features has leveraged off this. One approach has been to treat the HTML tree as a proxy for the database structure that the information was originally provided from and use this to locate records (Liu et al), (Zhai and Liu). Carlson and Shafer look for specific fields within records using a stacked classifier trained on character 3-grams and characteristics of text: all caps, all digits, title case, etc. Hao et al. use page layout, including size of features and location on the page, as well as text content, to improve the model accuracy.

Research Questions

Is this approach a significant benefit to blind users of the web? What are the design considerations for command line interfaces for blind users performing the most common tasks people do with web applications? Can current AI approaches achieve high enough accuracy to map user commands to web site controls to make TAB a viable approach? We decided to construct a prototype which would enable us to do some user testing as well as establish exactly what was needed from the AI. The prototype is built as an extension to the Firefox browser.

Prototype Interface Design

We decided on a Wizard of Oz (Kelley) approach to help us define the user interface (Wixon). The participant was provided with a PC running a chat application with text to speech enabled. This app was connected to a person acting as the wizard who was running the chat in one window and a browser opened to a grocery shopping site in another. The participant was asked to type whatever he felt was an appropriate dialog with an assistant to do his weekly grocery shopping. The wizard responded, and when appropriate, performed requested operations on the web site. The following design guidelines emerged from the experiment.

Accept synonyms: Both “add to cart” and “buy” appeared in the dialog, an interface might as well accept synonyms for common actions.

Accept typos: If the user types “serch”, it is pretty clear he meant “search”.

Flexibility in argument positioning and optional arguments: After hearing a list of product results, the user might type any of “buy 4^{th} item”, “buy 2, 4^{th} item”, buy fourth item, 2”. The meaning to the wizard of all of these was clear, the program should accept them as well.

List presentation: When a response to a user request provided a list of items, the most effective presentation was to provide a subset (5-7 items) of the total list with the ability for the user to skip to the next item if the beginning of the reading indicated it was not a desired item. Based on the user’s desire to specify an item in the list using an ordinal number, list items are prefixed with an alphabetic index character (a,b,c…) to make it easier for the user to specify the item he was referring to.

Ignore page boundaries: When a user is walking through a list, the number of entries of that list on any particular web page is irrelevant. TAB should fetch additional pages as needed.

Alonso observed that web site navigation often involves the user in walking down a tree of screens in order to complete an activity. For example, a user must transition from product search to cart in order to see the items in his cart. Sighted users are always aware of where they are in the tree, blind users need to keep this in their heads. He recommends that the depth of the tree be kept to a minimum.

In addition to these principles, the prototype was designed to allow minimum unique starting strings to identify commands, as well as single key shortcuts, and a help system. For the shopping domain, there are three major areas needing commands: product search and selection, cart management, and account management. For the prototype, only the first two were implemented, as this seemed adequate for the desired user testing. Table 1 shows the command set that was developed for these two areas. Multiple entries in the Commands column indicate synonyms for the command. “Index” is an ordinal number or single letter alphabetic index.

Table 2. TAB Commands for Shopping.

Commands	**Argument**	**Hotkey**	**Help Text**
search find	text	Ctrl + s	Search followed by text searches for that item in the product search screen
add to cart buy	[index] [qty]	Ctrl + a	Add to cart followed by the letter index of the products puts that item in the cart. You may add a quantity to specify more than one. Omitting the index letter adds the last read item to the cart.
filters refinements	[category]		If category is omitted, show list of filter categories. If present, show filters for that category
filter by set filter refine by	id		Filters the search results by the given filter name or index, as specified by id
filter off	id		Remove the specified filter
sort choices sort options			Provide a list of choices to sort by
sort by	id		Sort by specified id
cart show cart		Ctrl + k	Show list of items in cart

Commands	Argument	Hotkey	Help Text
change	index qty		Change the quantity of the item in the cart.
checkout		Ctrl + t	Start the checkout process
next	[num]	Ctrl + n	Reads the next group of items from the current list. Read number of items specified by num
previous	[num]	Ctrl + p	Reads the previous group of items from the current list. Read number of items specified by num
spacebar			Stop reading
down arrow			Read next item in list
up arrow			Read previous item in list

User Testing

This phase of testing was not intended to be a comprehensive validation of the concept, but rather a quick indication of whether the command line concept was likely to prove fruitful before investing time in developing the machine learning code.

Tim, a volunteer participant, was recruited to complete predefined tasks on each site, using both TAB and his usual browser (also Firefox) with JAWS as his screen reader. Tim normally used Peapod for his weekly shopping, so he was very familiar with the site. No learning of a new site was required. On the other hand, he was (naturally) unfamiliar with TAB. He was given a quick tutorial of the concept, and the help text for each command.

The following tables show the script and his commands in TAB and JAWS to fulfill the script. It does not show the speech response to each command. In the tables the symbol ? indicates that Tim had to listen to the response before his next keystroke. The symbol $ is shorthand for the Enter key. Numbers in parentheses indicate the command was repeated that many times.

Table 3. Script Fulfillment with TAB.

Script	**TAB keystrokes**	**TAB meaning**
Add 2 dozen large white eggs to cart	Ctrl + s white eggs$? Ctrl + a 2$	Search “white eggs”, read 1st five results Add 2 of first (index “a”) item to cart
Add 1 half gallon 2% milk to cart	Ctrl + s milk$? Ctrl + n? (5) Ctrl + a z$	Search “milk”, read 1st five results Get to 26th item, 5 at a time Add 26th (index “z”) item to cart
Add Jif crunchy peanut butter to cart	Ctrl + s jif peanut butter$? Ctrl + a c$	Search “jif…”, read 1st five results Add 1 of third item
Change milk to 2 half gallons	Cart$? ch b 2$	Read cart contents Change quantity of 2nd item to 2.

Table 4. Script Fulfillment with JAWS.

Script	**JAWS keystrokes**	**JAWS meaning**
Add 2 dozen large white eggs to cart	e?$white eggs$. h? (3) < down >? (2) <tab>? (2) $ h (3) <down>? (6) <tab>? (2) $	Find edit boxes, first one is search Jumping through headers, looking for “product list.” Jumping through list items, eggs is first Jumping through elements to find “add to cart” button Back to product list after page refresh Looking for “remove item from cart” button. Finding “add one more to cart button” and invoking
Add 1 half gallon 2% milk to cart	e$ milk$ h (3) <down> (26) <tab> (2) $	Going to search edit box Going to product list Find half gallon 2% milk Find and use “add to cart” button
Add Jif crunchy peanut butter to cart	e$Jif crunchy peanut butter $ h (3) <down> (2) <tab> (2) $	Going to search edit box Going to product list First item in products Find and use “add to cart” button

Script	JAWS keystrokes	JAWS meaning
Change milk to 2 half gallons	b? (2)$. h? (2) $ <tab>? (4) $	Searching for cart button, which is 2nd button on page. Jumping though headings to get to “Dairy”, the 2nd one Tabbing to get to the “add one more to cart” button

TAB: 210 seconds 64 keystrokes 9 times to listen for results.

JAWS 300 seconds 110 keystrokes 48 times to listen for results.

We considered the results encouraging, given that this was Tim’s first-time using TAB. We expect that with additional experience, his times with TAB would improve.

Tim was asked to repeat the experiment 18 months after it was initially done. During this time, he had switched to Instacart as his primary grocery shopping application. So, he was no longer as familiar with Peapod, now Stop and Shop. And they had made some changes to their web site. His Firefox/JAWS time increased to 40 minutes, demonstrating the difficulty blind users have with learning new web sites. Time with TAB remained the same.

Machine Learning Approach

Each user command requires that TAB identify one or more HTML elements to fulfill it. For example, “search” requires that TAB find the search text entry box element, and also be able to find the elements that constitute the product descriptions in response to the search. All of the shopping commands presented above require a total of 23 distinct elements to be found. Each requires a separate model.

There are two categories of capabilities to be developed: invoking actions, and extracting text. As a general rule, finding the actions is a somewhat simpler task, because the W3C accessibility guidelines require that control elements be labeled in some way for screen readers. Even web sites that are nonconforming usually have clear visual indications of action elements.

Text extraction is more difficult because there are no explicit labels for sections of a page.

The AI implementation defines a set of features that are used to train a model. A feature extractor is implemented that identifies these features within the rendered page. The approach we take focuses on features that are visual. When sighted users do a product search, they generally identify the results of the search by the fact that there are a number of similar boxes on the page. They are about the same size, include currency indicators and pictures, use larger fonts, etc. They are near the top, but not at the top.

Groupings of similar boxes in a regular grid show up not just in search but in a variety of other elements on the page such as menus, filters, shipping options etc. The feature extractor seeks to identify such groups. The group matching algorithm uses the determination of being an HTML sibling as a proxy for being in a grid. This is simpler to calculate and empirically works equally well. Some web sites break up groups into separate batches of siblings under different parents. The feature extractor processes found groups to join together ones that have similar characteristics.

Table 4 gives the criteria for determining that an HTML element is part of a group. The “Element Characteristic” column identifies the property of an element that is being considered for group membership. The “Evaluation criteria” column identifies the allowable difference between this element and the median value of that characteristic for the group as a whole, in order for this element to be considered part of the group.

Table 5. Characteristics To Identify Elements as Part of a Group.

Element Characteristic	**Evaluation Criteria**
HTML Tag	Tag must match most common tag of siblings
Complexity	Element must have at least 8 descendants.

Element Characteristic	Evaluation Criteria
Width	Element width must match within 10% of median height of elements in the group
Height	Element height must match within 70% of median height. No element in group can have height larger than screen height
Number of images in element	70%
Size of images in element	70%

Tolerance values were empirically determined by looking at a number of web sites. Results given below were not especially sensitive to variations in these values. Three sets of models were trained and tested, using the values above, and all of them modified by +-30%. There was no change to the results presented below.

The following features for each element fed to the Machine Learning algorithm are both individual and group characteristics extracted from the rendered page.

- Left and top coordinates, normalized to page size.
- Width, normalized to page width.
- Height and area, normalized to the screen.
- Top of group containing element.
- Number of elements in group containing element.
- Median area of elements in group containing element.
- Median area and number of images in group containing element.
- Font size and weight of element text
- Percentage of text characters that are alphabetic characters.
- Presence of currency symbol in text.
- Tag having a certain value (e.g., anchor, button, input). A separate feature for each value set to either 0 or 1 based on tag value.

- Element label having a certain value (e.g., “search”, “add”). A separate feature for each possible value. For the purposes of training features, “label” is broader than the HTML label element. It includes the text associated with an element as well as an aria-label that identifies the element.

ML Test Results

These features were used to train a Decision Tree model (Géron) on a set of training pages. The model was then used to identify the product description HTML element on untrained web pages. Training was done on three pages and testing was done on ten pages (including the training set pages) on each of ten web sites in the shopping domain: Amazon, Bestbuy, eBay, Etsy, Hobbyking, HomeDepot, StopandShop, Target, Walmart, and Wayfair, Table 5 shows the characteristics of the ten pages for each web site, in terms of the total number of HTML elements that are product descriptions, and the number that are not.

Table 6. Machine Learning Test Data.

Site	**Product Description Elements**	**Other Elements**	**% Product Description Elements**
Amazon	444	18,914	2.3%
Bestbuy	210	28,786	0.7%
eBay	2,399	36,976	6.5%
Etsy	640	9,916	6.5%
Hobbyking	360	9,710	3.7%
Homedepot	309	15,883	1.9%
Stopandshop	562	12,577	4.5%
Target	243	9,633	2.5%
Walmart	347	7,129	4.9%
Wayfair	456	16,426	2.8%

Each of the per web site models correctly recognized all elements across all ten pages of that site. As this is an appreciably better performance than is usually achieved with ML models, an analysis was done as to why such good results were achieved. Most web sites are designed to have a single, or small number of presentations for the results. Amazon, Home Depot, and Wayfair each have two different styles for presentation of product search results, all the rest have a single format. So, it is not surprising that a model trained on a small number of pages, can correctly predict other pages on the same site. A model was also built for finding the “add to cart” button in product descriptions and it also had perfect success.

Areas for Future Work

User testing with a single user clearly does not provide definitive evidence of the value of the TAB approach. It would appear worthwhile to perform efficiency testing with a larger set of blind computer users.

The ability to implement TAB depends primarily on the ability of AI to analyze a web page with enough accuracy to enable the mapping of a task-oriented command line interface over it. A given vertical domain will typically take several tens of models to support all of the activities necessary to complete a task. Work to date seems encouraging, but scaling up of model building is necessary to ensure that TAB is feasible.

Conclusion

A new approach to improve usability of web sites for blind users has been proposed. The approach provides command line instructions for completing common tasks and uses AI to locate the elements on a web page that are needed to fulfill tasks. This frees the user from the difficult job of locating the elements amidst the clutter that web sites typically provide. Work has been completed to demonstrate both the feasibility of the approach and the benefits to users. As is often the case with developments intended to improve accessibility for persons with disabilities, successful implementation of TAB would enable speech assistants to work on any TAB supported web site without the need for coding scripts for each site.

Acknowledgments

My thanks to Aaron Boxer, Tim Cumings, Eric Peters, and Bill Zimmer for their thoughtful discussions and help with this project. They constitute the “we” in this paper. My appreciation also to the Perkins School for the Blind, which initially identified the difficulty of blind users with web applications.

Works Cited

Alonso, Fernando, et al. "User-interface modelling for blind users." International Conference on Computers for Handicapped Persons. Springer, Berlin, Heidelberg, 2008. Systems, March 1984, 2:1, pp. 26–41.

Carlson, Andrew, and Charles Schafer. "Bootstrapping information extraction from semi-structured web pages." Joint European Conference on Machine Learning and Knowledge Discovery in Databases. Springer, Berlin, Heidelberg, 2008.

Géron, Aurélien. *Hands-on Machine Learning with Scikit-Learn, Keras and Tensorflow.* 2nd edition, O'Reilly, 2019, pp 175-185.

Hao, Qiang, et al. "From one tree to a forest: a unified solution for structured web data extraction." Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval. 2011.

Liu, Bing, Robert Grossman, and Yanhong Zhai. "Mining data records in web pages." Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. 2003.

Kelley, J. F., "An iterative design methodology for user-friendly natural language office information applications". ACM Transactions on Office Information Systems, March 1984, 2:1, pp. 26–41. [2]

Schneiderman, Ben, Catherine Plaisant. *Designing the User Interface: Strategies for Effective Human-Computer Interaction.* 5th edition, Addison-Wesley, 2010.

W3C. "Web Content Accessibility Guidelines (WCAG) 2.1". W3C, 5 June 2018, https://www.w3.org/TR/WCAG21/

Wixon, Dennis, et al. "Building a user-defined interface." Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 1983.

Zhai, Yanhong, and Bing Liu. "Web data extraction based on partial tree alignment." Proceedings of the 14th international conference on World Wide Web. 2005.*

2023
THE JOURNAL ON TECHNOLOGY AND PERSONS WITH DISABILITIES

Disability Bias & New Frontiers in Artificial Intelligence

Christopher W. Land
University of Michigan
cland@umich.edu

Abstract

Bias in artificial intelligence (AI) systems can cause discrimination against marginalized groups, including people with disabilities. Like other challenges with digital accessibility, this discrimination is most often unintentional and due to a lack of training and awareness of how to build inclusive systems. This paper has two main objectives: 1) provide an overview of AI systems and machine learning, including disability bias, for accessibility professionals and related non-development roles; and 2) discuss methods for building accessible AI systems inclusively to mitigate bias. An overview of current worldwide progress on establishing legal protection against AI bias is provided, with recommendations on strengthening laws to protect people with disabilities from discrimination by AI systems. When built accessibly, AI systems can promote fairness and enhance the lives of everyone, in unprecedented ways.

Keywords

Artificial Intelligence, Disability, Machine Learning, Bias, Big Data.

Introduction

Bias and discrimination in data science, artificial intelligence and machine learning are established problems that technologists have struggled to control since the inception of AI. The power of machine learning comes from pattern recognition within vast quantities of data. Using statistics, AI can reveal new patterns and associations that human developers might miss or lack the processing power to uncover. But these patterns are a best fit for the greatest average of the majority, and marginalized groups are often left out of the data. When they are, AI systems can be discriminatory (O'Neil, 2016).

The advent of new technology offers a range of possibilities that can be beneficial toward people with disabilities. For instance, robot assistants will help the elderly and people with disabilities with getting around and completing daily tasks independently. Advancing AI imaging and recognition tools will help nonvisual users understand images, video, and other visual information in the physical world. At the same time, biases occur in AI systems that inherently find patterns among the most typical majority, while minimizing outliers such as people in relative minority groups, raising concerns about systems discriminating against people with disabilities. Thus, laws in the United States will need to be strengthened to protect people from bias in AI systems. As with data privacy, other countries around the world are leading the way in legislating protection from the negative aspects of AI systems.

In the present paper, we will provide background on how artificial intelligence and machine learning work and explain how bias is introduced into AI systems if not proactively prevented. The approaches different nations are taking to reduce bias in AI systems and ensure that all citizens, regardless of disability status, are receiving fair and equal treatment within data science will be reviewed, and recommendations on strengthening legal protections for people with disabilities will be provided.

Discussion

Artificial Intelligence, Machine Learning and Big Data

To understand how artificial intelligence (AI) can inadvertently produce biased outcomes, it will first be helpful to review the basics of AI system development, and how it differs from traditional programming methods. Artificial intelligence is the broad science of emulating human intelligence with computers, including thinking, reasoning, learning, and making decisions. In the early attempts to create artificial intelligence, procedural programming was used. With procedural programing, human coders explicitly write the instructions for how the computer system will make decisions. This works well for simple directions, such as having your bank send you an email alert when your account balance is low. But the procedural approach to AI programming does not scale well. As programmers applied this explicit technique to hard, complex problems, such as diagnosing medical problems or predicting market behavior, the amount of code required became prohibitive. These systems were inaccurate and prone to mistakes (Domingos, 2015). Enter machine learning (ML).

With machine learning, human programmers do not create the decision engine – the computer creates the decision engine based on analyzing vast quantities of data (“big data”). Vast sets of data have become available in recent years as the price of computation and data storage has come down, and more and more human activity has moved to computer systems where data is recorded, such as online shopping and communicating via apps and social networks. This data can include user details and behavior, spoken and written language, images, financial information, science data, and more.

Machine learning algorithms process big data to discover patterns and relationships in the data without direct instructions from people, with amazing results. Machine learning has been powerfully successful in approaching the problems that procedural programming could not handle. For example, ubiquitous AI recommendation engines have become adept at predicting products we want and recommending media we will enjoy. ML is responsible for breakthroughs in image recognition, speech analysis, art creation, crime detection and more. The problem is that we do not have a clear understanding of how it works.

The Black Box

The decision rules created by machine learning are stored as multiple hidden layers of nodes, called a neural network. This neural network is referred to as a black box, because we cannot untangle the layers to understand the AI program's decision-making logic. We cannot look inside — it is not human readable. This has made bias visibility and analysis challenging (van Giffen et al., 2022).

Unlike ML-generated programs, procedural programs are written by humans and are human readable. Conditional logic determines the choices that a procedurally written program makes. For example, a code instruction written by a human coder to stop spam might use logic that detects spam using specific keywords (i.e., if the email subject included the words "free" and "easy", then send it to the junk folder).

To solve the spam email problem with ML, we feed the algorithm vast amounts of email labelled as spam and not spam. Based on this training data the machine will analyze this large set of data and determine the spam rules. The decision system created by machine learning will be highly accurate but is not human readable. This is called the Black Box problem since we are not able to look inside the ML-generated program to see how it is making decisions based on the analysis of the data. When there are problems with the results of ML-generated programs, such as biased or incorrect results, it is not possible “open the hood” of these programs to see what is happening.

Like the human brain, looking at the brain itself and individual neurons within cannot tell you what someone knows or thinks. The data is encoded in ways too complex for us to understand. To put it simply, we do not know how ML-generated programs work.

Increasing visibility into how AI decisions are made is the first step towards solving these problems of AI fairness. More recently, AI ethicists are advocating “Glass Box” machine learning models, which allow developers and researchers to better understand how AI systems make decisions to root out bias (Roelenga, 2022). The next frontier in AI transparency is explainable machine learning, in which the system is designed specifically to write human-readable rules and make its logic clear (Rai, 2021).

Bias and Discrimination

Discrimination against people with disabilities is a salient risk in artificial intelligence. As with age, gender, and race, machine learning systems can develop bias against disability status if the developers are not aware of the issue and do not design systems with inclusion intentionally. For example, AI tools used in the U.S. justice system have been used to inform judges’ decisions on which prisoners should be granted parole. Investigations have determined that these tools are racially biased, more frequently denying parole to African Americans. An AI hiring tool built by Amazon had to be scrapped when it showed a scoring bias against women. In both cases, bias built into the historical data against women and racial minorities will manifest in new artificial intelligence systems.

Race- and gender-related AI discrimination cases have made news headlines. To date, much of the work to eliminate bias within AI has focused on these efforts. At the same time, it is important to also consider how AI discriminates against people with disabilities and synchronize efforts to remedy these issues.

Without clear access to the decision-making system of the generated decision logic the issue of detecting and preventing bias in AI systems is challenging. Bias can be introduced in a number of ways, but the primary barrier to inclusive design of AI systems is bias in the training data. ML systems are only as good as the big data they are trained on. When the training data is not representative of the population, systems will work well for users represented by the data while malfunctioning for demographic groups not represented accurately in the data (Crawford, 2020).

If the data is historically accurate, but social behavior and cultural norms represented in the data is discriminatory, the system will be trained to perpetuate this discrimination. For example, an experimental job candidate approval system built by Amazon had to be dropped when it was shown to evaluate women more harshly than men, even when their qualifications were otherwise equal (Dastin, 2018). In 2015, Google Photos made headlines after the image recognition system mislabeled people with dark skin as gorillas (Barr, 2015). Presumably this system was not trained on a set of representative images. These cases and others received mainstream coverage, raising awareness on the issue of bias and the need for inclusion in the AI development process. But with emphasis on gender and race, are people with disabilities being included in these efforts?

Disability status bias is even more challenging to control for than age, gender or race bias – disability is not always shared or obvious, and the range of disability types makes this group very diverse. For example, someone with a mobility disability who requires more time to input their answers into the computer may fail a timed AI job screening test, while a blind user proficient with a screen reader may complete the test in the same time as a sighted mouse user. Conversely, if AI is used to measure a job candidate's eye contact during a video interview, the blind user may be penalized while the candidate with a mobility disability faces no obstacles. Some disabilities are visible while others are not.

Disability status may not be included in the training data, and therefore not considered by AI. We must recognize that the data sets used to train machines must be inclusive, and we must be aware of the potential bias so we can control for it. AI systems are deeply influenced by the data they are trained on. By their nature, machine learning systems optimize for norms. They are based on statistical frequency in data and therefore minimize outliers. Awareness and empathy are needed in the AI development community, since preventing disability bias requires vigilance and creative solutions.

In addition to discrimination inherent in current datasets, there are a host of other issues for people with disabilities related to the use of big data. For instance, big data collection can undermine health privacy protections and "out" people with disabilities without their consent. Even more troubling is that this data is owned by profit-driven companies and is unregulated. Data collected on user behavior via online traces can be compiled by social media and marketing companies using behavior to profile users with disabilities and market to or manipulate them. This could be harmless or even potentially beneficial to them, for example if this exposed the user to a new assistive technology for their disability. But marketers can also target users' vulnerabilities, such as feelings of insecurity (Zuboff, 2020).

Advertisers have been accused of using Facebook's platform to overtly discriminate against people with disabilities using collected data, and Facebook was penalized for this in 2019 by the Department of Housing and Urban Development (Booker, 2019). In choosing which groups would be delivered ads for housing properties, advertisers were given the option to omit certain groups, including people identified as showing interest in the topic of disability. HIPAA protects health-related information only as it applies to healthcare and related industries and does not apply to social media.

AI is used in other types of people analytics, including screening for jobs, college recruiting, and health and life insurance. If unchecked, these systems are likely to encourage discrimination against people with disabilities. For example, a user with disabilities taking an online job screening test may take longer completing the test using assistive technology. This person may be automatically screened out before being considered by a human interviewer. Globally, consider China's social credit system, which analyzes citizens' behavior using AI electronically and provides rewards and punishments based on what the state views as good behavior. The system expands tracking to all aspects of life. Get a ticket for playing loud music or post something unpatriotic online and you might lose the ability to buy plane tickets or rent an apartment in a nice neighborhood (Kobie, 2019). What happens when this type of program is not designed inclusively or is implemented in societies that stigmatize disability?

Controlling for Bias

Researchers have recently conducted a broad survey of general AI bias and compiled them into eight main types with recommendations on how to combat them specifically (van

Giffen et al., 2022). Two of these types represent the highest potential negative impact to people with disabilities and will be discussed here: social bias and representation bias.

Social bias (also known as historical or pre-existing bias) exists when the bias is or has historically been part of the culture or society and is therefore built into the new system. With representation bias (also known as population or selection bias) the bias is not present in society but is due to using training data that does not adequately represent the population. In both scenarios, the bias is built into the training data, and could result in the following scenario: When searching for an image using the word ‘family’ on a stock photo website, the traditional image of a mother, father and one or two children are most frequently shown, without photos of people with disabilities included. Because these families represent a minority percentage of families overall, even if the training data includes a representative sample, AI will tend to exclude outliers while establishing patterns in the data and will leave families including someone with a disability out of the results. This would be an example of social bias.

But, if images of families including someone with a disability are not included in the training data, this would be representational bias. In this case, the training data does not accurately represent the population, and the machine would have no way of including diverse families in its definition of ‘family’. If the desired outcome is to have families including people with disabilities represented in these search results, developers will need to make changes to the data and/or the algorithm to ensure diverse family types are included.

Broader approaches within the AI industry must also be adopted to prevent bias against people with disabilities. Inclusive design principles should be applied by hiring team members with disabilities and designing and testing with users with disabilities throughout the development process. Most discrimination in this area is unconscious or inadvertent – it is a common oversight to build systems for users like oneself. Having diversity within a team offers broader perspectives and activates people's natural empathy. Similar outcomes can be achieved by working with diverse users in brainstorming new system features and use cases, to reveal situations that may otherwise be overlooked. Testing inclusively with diverse data and real users will help address any unwanted outcomes before systems are deployed. Developers must be aware that bias is inherent in AI due to its propensity to find patterns that apply to the majority of the data while minimizing outliers, and continued vigilance against bias is therefore required. Companies should consider adopting ethical AI frameworks, including principles of fairness and inclusiveness to build thoughtful and responsible design into the development culture. Microsoft's AI principles (Microsoft) offer a great example.

Education on AI bias and discrimination must be expanded so AI developers design systems to be inclusive from the start. Tools such as IBM's AI Fairness 360 toolkit are available to developers to assist in mitigating bias in training data and ML outcomes. More work in this area is needed.

Legal Protection

Some governments around the world are taking steps to mandate fairness and protect citizens from bias in AI systems. Brazil's House of Representatives has passed the Brazilian Artificial Intelligence Bill (Bill No. 21/2020) which is now in review by the Senate. This bill establishes fundamental principles in AI development, specifically including human rights and respect for diversity and inclusion. If signed into law, discrimination by AI systems in Brazil will be explicitly illegal (Kujawski et al., 2022).

China's Ministry of Science and Technology has released governing principles for AI, including fairness for all people (CDIC, 2019). While these positive high-level guiding principles are a step in the right direction, fairness in this guidance may not be sufficiently defined to provide adequate protection for minority groups (Wu et al., 2020). Similarly, efforts are underway in the European Union and U.S. to draft and release guidelines for AI development promoting fairness and non-discrimination against groups, but these efforts fall short of legal regulations at this time (Madiega, 2019). Eighteen U.S. states have introduced legislation related to negative potential outcomes to citizens from AI. So far, four states (Colorado, Illinois, Vermont, and Washington) have passed AI governance legislation—these approved bills are related to forming committees to investigate the issue of AI fairness (NCSL, 2022). There have been calls for regulation around the world and guiding ethical frameworks published, but little in the way of enforceable regulation on AI fairness has been established to date.

To protect people with disabilities from the potential risks of AI discrimination in the U.S., federal legal protection needs to be expanded, including the ADA (Americans with Disabilities Act) and HIPAA (Health Insurance Portability and Accountability Act). The ADA was enacted before these technologies were prevalent, and the specific reference to non-discrimination in “places of public accommodation” as worded in the ADA has been challenging to enforce in digital spaces. Explicit clarification protecting people from discrimination by AI systems is needed. HIPAA protects people’s health privacy only within healthcare systems and healthcare-related industries. It cannot be applied to industries outside of healthcare, such as marketing companies or social networks. Health privacy should be protected across all industries.

Conclusion

The power and ubiquity of artificial intelligence continues to grow rapidly. Used with care these technologies hold the potential to enhance life for everyone, including people with disabilities. With empathy and awareness, we can and must ensure that traditionally marginalized groups are treated with increased fairness and are not left behind.

While providing breakthroughs in the power and accuracy of AI systems, the Black Box problem created by neural networks presents danger when we do not fully understand how our programs are making decisions. Care must be taken to ensure the data that systems are trained on is intentionally balanced for fair outcomes, and systems must be continually monitored for discriminatory results. Developments in glass box and explainable AI systems will help mitigate this problem. Stronger legislation must be passed to ensure fair and equal treatment for all people by AI tools.

The emerging technologies of machine learning, big data and artificial intelligence open up wonderful opportunities in advancing accessibility. The potential for increased discrimination is also a risk, which can hopefully be mitigated through education and expanded legal protection for people with disabilities. Further specific research into AI bias against people with disabilities is needed. Increasing advocacy and awareness on AI fairness is critical toward supporting equity in our technology driven world.

Works Cited

Barr, Alistair. "Google Mistakenly Tags Black People as 'Gorillas,' Showing Limits of Algorithms." *The Wall Street Journal*, 1 July 2015, https://www.wsj.com/articles/BL-DGB-42522. Accessed 18 January 2021.

Booker, Brakkton. "Housing Department Slaps Facebook with Discrimination Charge." *NPR*, 29 March 2019, https://www.npr.org/2019/03/28/707614254/hud-slaps-facebook-with-housing-discrimination-charge. Accessed 10 February 2021.

China Daily Information Co (CDIC). "Governance Principles for the New Generation Artificial Intelligence – Developing Responsible Artificial Intelligence." *China Daily*, 17 May 2019, https://www.chinadaily.com.cn/a/201906/17/WS5d07486ba3103dbf14328ab7.html. Accessed 5 November 2022.

Crawford, Kate. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence.* Yale University Press, 2020.

Dastin, Jeffrey. "Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women." *Reuters*, 10 October 2018, https://reut.rs/2Od9fPr. Accessed 15 October 2022.

Domingos, Pedro. *The Master Algorithm.* Basic Books, 2015.

Kobie, Nicole. "The Complicated Truth About China's Social Credit System." *Wired*, 6 July 2019, https://www.wired.co.uk/article/china-social-credit-system-explained Accessed 2 October 2019

Kujawski, Fabio F., et al. “Upcoming Brazilian Artificial Intelligence Act: What Businesses Need to Know.” *Lexology*, 13 May 2022. https://www.lexology.com/commentary/tech-data-telecoms-media/brazil/mattos-filho-veiga-filho-marrey-jr-e-quiroga-advogados/upcoming-brazilian-artificial-intelligence-act-what-businesses-need-to-know Accessed 5 October 2022.

Madiega, Tambiama. *EU Guidelines on Ethics in Artificial Intelligence: Context and Implementation*. European Parliamentary Research Service, PE 640.163, September 2019.

Microsoft. “Our Approach to Responsible AI at Microsoft”. https://www.microsoft.com/en-us/ai/our-approach. Accessed 5 November 2022.

National Conference of State Legislatures (NCSL). “Legislation Related to Artificial Intelligence.” 26 August 2022, https://www.ncsl.org/research/telecommunications-and-information-technology/2020-legislation-related-to-artificial-intelligence.aspx. Accessed 15 October 2022.

O’Neil, Cathy. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown, 2016.

Rai, Arun. “Explainable AI: From Black Box to Glass Box.” *Journal of the Academy of Marketing Science*. 48, 137–141 (2020). https://doi.org/10.1007/s11747-019-00710-5.

Roelenga, Bas. “Think Outside the ‘Black’ Box.” *Towards Data Science*, 21 April 2021, https://towardsdatascience.com/think-outside-the-black-box-7e6c95bd2234. Accessed 1 October 2022.

Van Giffen, Benjamin, Dennis Herhausen and Tobias Fahse. "Overcoming the Pitfalls and Perils of Algorithms: A Classification of Machine Learning Biases and Mitigation Methods." *Journal of Business Research*, vol 144, 2022, pp. 93-106. https://doi.org/10.1016/j.jbusres.2022.01.076.

Wu, Wenjun, Tiejun Huang, and Ke Gong. "Ethical Principles and Governance Technology Development of AI in China." *Engineering*, vol 6, no. 3, 2020, pp 302-309. https://doi.org/10.1016/j.eng.2019.12.015.

Zajko, Mike. "Artificial intelligence, algorithms, and social inequality: Sociological contributions to contemporary debates." *Sociology Compass*, 16(3) March 2022, e12962. https://doi.org/10.1111/soc4.12962.

Zuboff, Shoshana. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs, 2020.

Multilayer Information Obligation, and Why We Need It

Maksymilian M. Kuźmicz
Stockholm University
maksymilian.kuzmicz@juridicum.su.se

Abstract

Information obligation is a key element of consumer protection, particularly in the context of innovative products. This is especially important when it comes to assistive technologies (AT) for senior citizens and people with disabilities. Being informed about the functioning of the system, potential risks, but also one's rights is crucial for building trust in technologies. Information obligation in its current form is often criticised as ineffective. An introduction of multilayer information obligation (MIO) may be an innovative and promising improvement.

This paper aims to present the concept of MIO and identify the benefits of that solution. Firstly, three main problems with information obligation in its current form are acknowledged. Secondly, MIO is presented as a solution to all these problems. The key idea of a proposed mechanism is to introduce three layers of information obligations, each more detailed than the previous one. The content, form, and extension of the obligation of each layer are discussed. Thirdly, the paper examines the main benefits and risks of MIO. Finally, the possible implementation, either through legal requirements or voluntary standards adopted by the industry, is discussed.

Keywords

Information obligation, consumer protection, technological acceptance, data protection.

Introduction

Information obligation is a key element of consumer protection, particularly in the context of innovative products. It helps to mitigate the negative effects of an informational imbalance between sellers and consumers and contributes to the increased social acceptance of products (Kuźmicz, 261-262). This is especially important when it comes to assistive technologies (AT) for senior citizens and people with disabilities. AT help users deal with everyday activities, and often process their data. Being informed about the functioning of the system, potential risks, but also one's rights is crucial for building trust in technologies (Guner and Acaturk, 311. Davis, 320). Information obligation in its current form is often criticised as ineffective. An introduction of multilayer information obligation (MIO) may be an innovative and promising improvement.

This paper aims to present the concept of MIO and identify the benefits of that solution. Firstly, three main problems with information obligation in its current form will be acknowledged. Secondly, MIO will be presented as a solution to all these problems. The key idea of a proposed solution is to introduce three layers of information obligations, each more detailed than the previous one. The content, form, and extension of the obligation of each layer will be discussed. Thirdly, the paper examines the main benefits and risks of the proposed solution. Finally, I will address the question of implementation: either through legal requirements or through voluntary standards adopted by the industry. It is important to notice that this paper is written primarily from the legal perspective, arguing for regulatory measures, evaluating their effectiveness and highlighting potential consequences for all stakeholders, with a special focus on the protection of the fundamental rights of users. Social and technical aspects are identified and acknowledged as vital, but discussing them in detail will not be feasible, and goes beyond my scientific competencies.

This research does not deal with any particular AT but is more general. There is a plethora of AT systems that differ from each other significantly. These differences may be addressed in further, more detailed studies. In this paper, I follow the definition of AT and assistive products by WHO's Global Cooperation on Assistive Technology (GATE). More advanced AT that employ AI, video or audio are more challenging, also in the context of information obligation (Smith). Therefore, this study is mostly focused on them.

Discussion

Critique of Information Obligation in its Current Shape

Information obligation is a valuable tool for consumer protection, but its existing form is not effective enough (Kuźmicz 266-267). Multiple authors acknowledge various shortcomings of how information obligation is shaped now (Apan and Miff. Sharma. Wikström. Janssen). There are three most recurring and impactful problems that lower the efficiency of information obligation:

1. Users' rational ignorance.
2. Ineffective form of informational documents.
3. The tension between simplicity and comprehensiveness of information.

Rational ignorance means that people tend to ignore information when the expected cost of reading them (time and effort) is too high when compared with the potential benefit (Paredes). It happens often with informational documents. They are rich in diverse types of information and detailed, provide more data than consumer expects and can read. As a result, users are overwhelmed by the amount of delivered information (Bowler and Nicholson, 381). Potential users of AT, especially people with disabilities and senior citizens, may feel overwhelmed with information which can negatively impact their acceptance of AT (Ziefle et al, 590).

The forms of information documents are not effective. Studies show that informational documents are often too lengthy, use professional jargon that is not clear to most people, and lack clear structure and layout (London Economics and Ipsos. OECD. Klepper et al.). For senior citizens and people with disabilities, the size and font of letters, spacing, and colour of the background may create an additional challenge. Moreover, because of differences in the form of informational documents of similar products, consumers struggle to compare products. In other words, because of that specific form, information obligation does not achieve its purpose.

There is a tension between simplicity and comprehensiveness that are both expected from informational documents (London Economics and Ipsos. Van Dyck, 78. Colaert, 231). On the one hand, the theory of economics and the law require that consumers receive comprehensive information, which is a wish of many consumers too (Grundman et al. Beales et al). On the other hand, people want simple information that does not require professional knowledge to be understood (Guner and Acaturk). In that spirit, many European documents relevant to AT require providing information in an accessible way (GDPR, Article 12(1), 31(2). AI Act, Article 13 (2)).

All three presented problems can be potentially solved by introducing a more effective form of informational documents. It directly answers the second problem but can contribute also to lowering the tension between simplicity and comprehensiveness, and limiting the overflow of information for users.

Multilayer Information Obligation

A General Idea

MIO aims to solve the aforementioned problems by introducing three layers of information obligations, each more detailed than the previous one:

1. Layer 1 (L1) – key features of the product, i.e. privacy intrusiveness, use of AI.
2. Layer 2 (L2) – performance information, i.e., where data is processed, use of profiling.
3. Layer 3 (L3) – detailed description of the product, i.e., applied AI techniques, applied Privacy Enhancing Technologies (PET).

Providers of AT are obliged to provide various information. Some of them are common for all products, but some are specific for medical devices, products employing AI, or solutions processing personal data. This information differs in its specificity and level of detail. Some of them are very general, i.e., that AT process data related to the health condition of the user. Delivered information may have a different value for users. This value may depend on personal preferences. Some information is perceived as key ones in the context of similar systems. MIO proposed in this paper is focused on advanced AT that applies various sensors, including cameras, and technologies which may be classified as AI.

The first layer of MIO would provide consumers with the most general information about the key features of the product. For AT, these key features may be the level of privacy intrusiveness, use of AI, or transfer of personal data to a third party or abroad. Such general information may be presented in the form of a graphic or numeric label, i.e., privacy intrusiveness could be expressed in a form of an open, half-open, or closed eye, or an eye in various colours, from green (non-intrusive) to red (strongly-intrusive), or by privacy intrusiveness index on the scale 1-10. That would allow consumers to pre-screen products, and reject options that have some unacceptable features, e.g., high privacy intrusion. A graphic form may be a challenge for visually impaired people, therefore numeric labels shall be favoured. However, the number of AT allowing visually impaired people to read graphics is constantly increasing (Ramôa, 185-186).

Information on L2 should be system performance information. This means information about what data is used (location, health-related, biometric, performance etc.), if profiling is employed, where data is processed (locally, on the provider's servers, in the cloud) etc. This information is more detailed and should allow consumers to understand how the product works. L2 information should be presented in a simple form: tables or "yes" or "no" questions, e.g., "Profiling – yes/no. Data processed on the device – yes/no).

The last level of information obligation would contain a detailed description of the product. L3 information describes the applied technologies (AI techniques, security mechanisms, PET) and logic of the system, and should allow consumers to understand the workings of the system. Layer 3 may be further divided into two subgroups: regular, and professional. The regular informational document would be targeted at non-professionals, people with average knowledge. Such a document shall be written in narrative form, including examples (brief scenarios), and avoid jargon. The professional informational document should be more detailed, and inform users about the applied technical solutions, but without details that could put at risk the know-how of the provider.

Content and Form

The exact content and form of information of each layer should be determined based on studies. There is a lack of behavioural research on the perception of information in the context of AT, but also processing personal data or AI solutions. Regarding what information should be delivered on which level, it is crucial to find what are the expectations of users. Presumably, privacy-intrusiveness and application of AI may be L1 information for most users (Mittelstadt, 163). Measurement of the key features of AT is another challenge. Over eighty different privacy metrics have been proposed by scholars, and the discussion on how to evaluate or merge them is outgoing (Wang et al. Wagner and Eckhoff,). This paper does not aim to solve that problem. The proper measurement, taking into account fundamental rights law, cultural and social context, and philosophical reflections, should be prepared in the course of interdisciplinary research, and tested empirically (Martin and Nissenbaum).

Similarly, the form of presentation of information at each layer should be determined in detail following behavioural research. Some preliminary ideas may be taken from the research on information obligation in the financial sector. Wild-scale research conducted for EU institutions after the financial crisis of 2008 reveals people's preference for graphical forms, numerical scales, tables, and narrative language (Lusardi et al, 301). Therefore, these forms are suggested for information obligation for L1 and L2. Consumers point out the aforementioned forms as the easiest to comprehend and stated that they skip many parts of the text but almost always "read" graphic or numeric labels (London Economics and Ipsos). An informational document of L3 may have a form of a text. It should not be too long, at least in its regular version, as users tend to o lose focus after a few lines if they know that the document is long (OECD). Crucially, the forms of informational documents on each layer should be harmonised to enhance the comparability of products. A clear structure, similar to various products, makes information easier to be processed and used by consumers (Klapper et al).

Delivery of Information

Providers of AT should be obliged to provide consumers with L1 and L2 information, while L3 information should be delivered at the consumer's request. Otherwise, consumers could be overwhelmed with information even more than they are now, as they would receive information in three various forms (Parades. London Economics and Ipsos). On a practical level, for products with a physical component, L1 and L2 information may be presented on the box of the device, and for L3 on the website of the provider. For software solutions, L1 and L2 information can be provided in a form of a pop-up window or a bar on the website, while level 3 information upon clicking a link.

There is a potential problem with the proposed difference in the delivery of information. Providers of AT may be legally required to provide some information. In the EU, in case of a dispute whether the information was delivered is on the provider of AT (Orange România SA v ANSPDCP. Szpunar). It is not clear whether a QR code on the box or link on the website would be considered by the Court of Justice of the European Union (CJEU) as sufficient measures to fulfil the information obligation. Van Calster notices, that “if click-wrapping makes it possible to print and save the text of those terms and conditions before the conclusion of the contract, then it can be considered a communication by electronic means which provides a durable record of the agreement” (122). In El Majdoub v CarsOnTheWeb, CJEU rejected the claim that a pop-up window is required to fulfil information obligation in the context of a normal sale contract. However, in the context of data protection or information on medical devices, which many AT are, CJEU may adopt a different position. If it does so in its future judgements, all mandatory information will have to be moved to L1 or L2.

Main Benefits and Risks

Multilayer information obligation would be beneficial in many ways. Firstly, consumers would receive more transparent and understandable information, due to the employed forms. That is especially essential for users of AT who are facing additional challenges. Secondly, consumers would be immediately informed about the key features of the product (L1 information), with a guarantee of access to more detailed information. Thirdly, harmonisation of the form in which information obligation is performed shall make it easier for consumers to compare products. As a result, senior citizens and people with disabilities may choose a technological solution that fits best their needs. Finally, the standardisation of information documents will allow companies to fulfil their information obligation more effectively (Kuźmicz, 265).

The proposed solution is not free of risk. Multiple layers of information obligation may increase compliance costs (Enriques and Gilotta, 539). However, harmonisation and simplification of forms should bring compliance costs down in the long run. The other risk is that providers will try to manipulate data about the features of their products to fit in a better category, especially when it comes to L1 and L2 information, and in consequence present AT as more attractive. That risk cannot be completely avoided but can be mitigated by careful design of the forms. The construction of the forms, and especially of graphic labels, is a big challenge. It requires carefully choosing the key features of AT, but also proposing scale and measurement criteria. For that reason, it is of utmost importance that the legislature works closely with interdisciplinary research teams, and with organisations of civil society.

Way of Implementation of the Multilayer Information Obligation

Implementation of the idea of MIO is a challenge on its own. One of the main questions is whether it should happen through new regulations or voluntarily adopted standards. The advantages and disadvantages of each in different contexts have been thoroughly discussed by scholars. (Lipson. Abbott and Snidal 2000, 2004. Koremenos et al. Trubek et al). It is worth noticing that the implementation of MIO should not need new legislation. A text form of the informational document is rather a custom of AT providers than a legal requirement. Many legal acts do not prohibit graphic or numeric labels but even explicitly allow them. For example, in the EU data protection law, recital 60 to GDPR explicitly states that the required by law information “may be provided in combination with standardised icons to give in an easily visible, intelligible and clearly legible manner, a meaningful overview”.

In the ideal scenario, MIO should be introduced primarily through the voluntarily adopted standards (soft law). Cooperation between providers of AT and scientists can produce effective solutions, introduced in the form of guidelines, codes of best practices, or technical standards. There are at least three potential benefits of this method of implementation that should be pointed out. First, research show that voluntarily adopted methods are more successful in achieving ambitious goals (Skjærseth et al, 118. de Hert and Papakonstantinoua, 10-11). Second, soft law mechanisms provide greater flexibility and tend to pay more attention to the recommendations of scientists (Skjærseth et al, 118-119. Schaffer and Pollack, 7-8). Third, if AT providers adopt MIO on their own, they can see it as their commitment, and not as an additional burden. The history of the Nutri-Score label can be an inspirational example of a science-based, non-mandatory graphic information system (Julia et al).

However, soft law instruments may be insufficient. AT providers may not be convinced to implement a system that in a brief time will generate additional costs. What is even more fundamental, MIO's goal is to provide people with disabilities and senior citizens with more accessible information, and by that support them in using AT. Legal regulations significantly enhance the capacity for enforcement (Abbott and Snidal 2000, 427). Shaffer and Pollack notice, that it increases the credibility of rules (6). This is crucial in the context of AT. Users must know that they are provided with adequate, sufficient information, allowing them to make an informed choice (Sitnik). Therefore, the legislature shall consider drafting a law introducing MIO.

Conclusions

Reflections on the concept of MIO show that this idea has the potential to help consumers, including users of AT, by providing them with information. This information would not differ much in terms of content from the current informational documents, but would be more understandable, easier to process, and will be harmonised within similar products. MIO can benefit both providers and users, and establish a new standard in information documents. Current information imbalance results in an imbalance of power and lower acceptance of technology, which stops many people from using AT. Multilayer information obligation may change it and allow more people to benefit from technological innovations.

However, there are still many issues that require further research:

1. There is a need for behavioural, sociological, and legal studies to decide what information should be delivered on each layer.
2. Multidisciplinary research should formulate appropriate measurements of key features of AT, propose new forms for the informational documents, and test them empirically.

3. Key concepts used in the discussion over AT must be operationalised so they can be used at L1 as features of systems. Terms like privacy, security, and data protection, should be defined synthetically, taking into consideration legal, social, philosophical, technological, and economic contributions. At the same time, these definitions should recognise some gradation or spectrum that would be used at L1 and L2.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 861091. This document reflects the views only of the authors, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

The author thanks Ioannis Tsitsovits (KU Leuven) and Luca Love (Karolinska Institutet) for reading the previous version of this paper, and for their valuable comments.

Works Cited

Court of Justice of the European Union. *El Majdoub v CarsOnTheWeb*. 21 May 2015.

Court of Justice of the European Union. *Orange Romania SA v Autoritatea Naţională de Supraveghere a Prelucrării Datelor cu Caracter Personal (ANSPDCP)*. 11 November 2020.

European Union, Council of the European Union and European Parliament. "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)". *Official Journal of the European Union 119,* European Union, 2016, pp. 1-88. https://eur-lex.europa.eu/eli/reg/2016/679/oj.

Abbott, Kenneth W., and Duncan Snidal. "Hard and Soft Law in International Governance." *International Organization*, vol. 54, no. 3, 2000, pp. 421–456., doi:10.1162/002081800551280.

Abbott, Kenneth, and Duncan Snidal. "Pathways to Cooperation." *The Impact of International Law on International Cooperation: Theoretical Perspectives*, edited by Benveniśtî Eyāl and Moshe Hirsh, Cambridge Univ. Press, 2011, pp. 50–84.

Apan, Rodica-Diana, and Angela Miff. "Valences of the Information Obligation under the Package Travel Contract in the Age of Globalisation Vs 'The Travel of Dilettantes'?" *Fiat Iustitia*, vol. 1, Jan. 2019, pp. 24–42.

Beales, Howard, et al. "The Efficient Regulation of Consumer Information." *The Journal of Law and Economics*, vol. 24, no. 3, 1981, pp. 491–539., doi:10.1086/466997.

Calster, Geert van. *European Private International Law*. Hart, 2016.

Colaert, Verlee. “Building Blocks of Investor Protection: Ever expanding regulation tightens its grip.” *Journal of European Consumer and Market Law*, 2017, p. 229.

Davis, Fred D. “Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology.” *MIS Quarterly*, vol. 13, no. 3, 1989, p. 319., https://doi.org/10.2307/249008.

De Hert, Paul, and Vagelis Papakonstantinou. “Framing Big Data in the Council of Europe and the EU Data Protection Law Systems: Adding ‘Should’ to ‘Must’ via Soft Law to Address More than Only Individual Harms.” *Computer Law & Security Review*, vol. 40, 2021, p. 105496., doi:10.1016/j.clsr.2020.105496.

Enriques, Luca, and Sergio Gilotta. “Disclosure and Financial Market Regulation.” *SSRN Electronic Journal*, 2014, doi:10.2139/ssrn.2423768.

Grundmann, Stefan, et al. “Party Autonomy and the Role of Information in the Internal Market – an Overview.” *Party Autonomy and the Role of Information in the Internal Market*, by Stefan Grundmann et al., Walter De Gruyter, 2001, pp. 3–38.

Guner, Hacer, and Cengiz Acarturk. “The Use and Acceptance of ICT by Senior Citizens: A Comparison of Technology Acceptance Model (TAM) for Elderly and Young Adults.” *Universal Access in the Information Society*, vol. 19, no. 2, 2018, pp. 311–330., https://doi.org/10.1007/s10209-018-0642-4.

Guner, Hacer, and Cengiz Acarturk. “The Use and Acceptance of ICT by Senior Citizens: A Comparison of Technology Acceptance Model (TAM) for Elderly and Young Adults.” *Universal Access in the Information Society*, vol. 19, no. 2, 2018, pp. 311–330., doi:10.1007/s10209-018-0642-4.

Janssen, André. "He Update Obligation for Smart Products – Time Period for the Update Obligation and Failure to Install the Update." *Smart Products*, edited by Sebastian Lohsse et al., Nomos, 2022, pp. 91–108.

Julia, Chantal, et al. "Front-of-Pack Nutri-Score Labelling in France: An Evidence-Based Policy." *The Lancet Public Health*, vol. 3, no. 4, 2018, doi:10.1016/s2468-2667(18)30009-4.

Koremenos, Barbara, et al. "The Rational Design of International Institutions." *International Organization*, vol. 55, no. 4, 2001, pp. 761–799., doi:10.1162/002081801317193592.

Kuźmicz, Maksymilian M. "Information obligation as a balancing tool in the context of Active and Assisted Living." *ICCHP-AAATE 2022 Open Access Compendium „Assistive Technology, Accessibility and (e)Inclusion"*, edited by Petz, Hoogerwerf, Mavrou, Association ICCHP, 2022, pp. 260-269.

Lipson, Charles. "Why Are Some International Agreements Informal?" *International Organization*, vol. 45, no. 4, 1991, pp. 495–538., doi:10.1017/s0020818300033191.

London Economics and Ipsos. Consumer testing study of the possible new format and content for retail disclosures of packaged retail and insurance-based investment products – Final Report. 2015.

Lusardi, Annamaria, et al. "Visual Tools and Narratives: New Ways to Improve Financial Literacy." 2014, https://doi.org/10.3386/w20229.

Martin, Kirsten E., and Helen Nissenbaum. "Measuring Privacy: Using Context to Expose Confounding Variables." *SSRN Electronic Journal*, 2015, doi:10.2139/ssrn.2709584.

Mittelstadt, Brent. “Ethics of the Health-Related Internet of Things: A Narrative Review.” *Ethics and Information Technology*, vol. 19, no. 3, 2017, pp. 157–175., doi:10.1007/s10676-017-9426-4.

Paredes, Troy A. “Blinded by the light: information overload and its consequences for securities regulation.” *Washington University Law Quarterly*, vol. 81, 2003, p. 417.

Ramôa, Gaspar. “Classification of 2D Refreshable Tactile User Interfaces.” *ICCHP-AAATE 2022 Open Access Compendium, Assistive Technology, Accessibility and (e)Inclusion",* edited by Petz, Hoogerwerf, Mavrou, Association ICCHP, 2022, pp. 185-192.

Shaffer, Gregory. “Hard and Soft Law: What Have We Learned?” *International Law and International Relations: Insights from Interdisciplinary Scholarship,* by Mark A. Pollack and Jeffrey L. Dunoff, Cambridge University Press, 2013.

Sharma, Sanjay. *Data Privacy and GDPR Handbook*, by Sanjay Sharma, Wiley, 2020, pp. 193–232.

Sitnik, Piotr. “Intricacies of Significant Imbalance as the Cornerstone of Protection against Unfair Terms in Consumer Contracts under EU Law.” *Studia Iuridica*, vol. 82, 2020, pp. 227–262., doi:10.5604/01.3001.0013.9789.

Skjærseth, Jon Birger, et al. “Soft Law, Hard Law, and Effective Implementation of International Environmental Norms.” *Global Environmental Politics*, vol. 6, no. 3, 2006, pp. 104–120., doi:10.1162/glep.2006.6.3.104.

Smith, Roger O., et al. “Assistive Technology Products: A Position Paper from the First Global Research, Innovation, and Education on Assistive Technology (Great) Summit.” *Disability and Rehabilitation: Assistive Technology*, vol. 13, no. 5, 2018, pp. 473–485., doi:10.1080/17483107.2018.1473895.

Szpunar, Maciej. *AG Szpunar Opinion in Planet49*. 2019.

Trubek, David, et al. "'Soft Law,' 'Hard Law,' and European Integration: Toward a Theory of Hybridity." *Law and New Governance in the EU and the US*, by Búrca Gráinne De and Joanne Scott, Hart, 2006, pp. 65–94.

Van Dyck, Tom. *De geharmoniseerde prospectusplicht*. Die Keure, 2010.

Wagner, Isabel, and David Eckhoff. "Technical Privacy Metrics." *ACM Computing Surveys*, vol. 51, no. 3, 2019, pp. 1–38., doi:10.1145/3168389.

Wang, Yong, et al. "Social Network Privacy Measurement and Simulation." *2014 International Conference on Computing, Networking and Communications (ICNC)*, 2014, doi:10.1109/iccnc.2014.6785440.

Wikström, Solveig. "Bringing Consumer Information Systems down to Earth. Experiences from a Swedish Experiment." *Journal of Consumer Policy*, vol. 7, no. 1, 1984, pp. 13–26., doi:10.1007/bf00380288.

Ziefle, Martina Ziefle, et al. "Perceived Usefulness of Assistive Technologies and Electronic Services for Ambient Assisted Living." *Proceedings of the 5th International ICST Conference on Pervasive Computing Technologies for Healthcare*, 2011, doi:10.4108/icst.pervasivehealth.2011.246044.

THE JOURNAL ON TECHNOLOGY AND PERSONS WITH DISABILITIES

European Digital Rights – Human Rights for a Digital Age?

Maksymilian M. Kuźmicz
Stockholm University
maksymilian.kuzmicz@juridicum.su.se

Abstract

Technological innovations in assistive technologies create great opportunities, but they come with big challenges too. One of them is the proper protection of the rights of all stakeholders, but especially those belonging to vulnerable groups, i.e., senior citizens and people with disabilities. It may be beneficial to draft a new catalogue of fundamental rights. This paper examines if the European Declaration on Digital Rights may be the first step in that direction. For that purpose, Humans Rights (HR) are discussed as supreme principles of legal systems, from the perspectives of philosophy and law. Secondly, two key features of the European HR law system are identified. Thirdly, legal aspects of the Declaration are examined, and compared with HR law. The legal analysis of the Declaration is divided into four steps: the legal position of the Declaration, material analysis of the proposed Digital Rights, enforcement mechanisms, and comparison with the European Convention on Human rights. In conclusion, the importance of the Declaration and the potential of European Digital Rights are acknowledged, and the most needed amendments are proposed.

Keywords

Human Rights, Digital Rights, EU law, data protection, AI.

Introduction

Technological innovations in assistive technologies (AT) create great opportunities, but they come with big challenges too. While they allow people to be more autonomous and enjoy their lives, they may make people feel more dependent on technology. People often do not understand the technologies they use, which makes users feel insecure. At the same time, there are a variety of interests connected with AT. Balancing these interests and protecting the rights of all stakeholders is a challenge for the courts, legislature, and legal scholars.

The need for effective protection of rights and interests is especially important in the case of persons with disabilities and older citizens. First, people with disabilities face challenges and have specific needs. Because of that, there is a duty to provide them with enhanced protection (Limante, 30). Secondly, people with disabilities tend to feel vulnerable more often than others. While AT have the potential to elevate the users' quality of life, they also make them dependent on the system they use (Ake-Kob et al). Finally, AAL solutions deal with personal data. In many cases, it is data concerning health conditions, habits, and relations. Processing that data is considered by many as privacy intrusive and creates an imbalance between the user and the provider of the technology (Solve, Calo). Such deprivation of privacy could result in even increased feeling of vulnerability.

One solution to these challenges might be to draft a new catalogue of fundamental rights. That idea has been researched by many scholars (Mathiesen, Custers). The European Declaration on Digital Rights and Principles (Declaration), proposed by the European Commission, is the first attempt to provide a full catalogue of probable new fundamental rights. There is no research on the Declaration and its potential significance. While it is still in its early stage, it is vital to initiate a discussion over it, and locate it properly in the context of ongoing discourse. This paper aims to examine whether the rights listed in Declaration have the potential to become new Human Rights (HR), especially in the context of AT. For that purpose, HR will be discussed as supreme principles of legal systems, from the perspectives of philosophy and law. Secondly, some key features of the HR law system will be identified. Thirdly, legal aspects of the Declaration will be examined, and compared with HR law.

Discussion

Human Rights as the Supreme Legal Order

In many jurisdictions, including the American and European ones, Human Rights occupy a supreme position in the legal system. Out of many aspects of that supremacy, I propose to distinguish three: philosophical, jurisprudential, and institutional.

Human Rights law stems from the philosophical belief that some fundamental rights are common to all human beings (Hayden. Freeman). All people possess HR simply by virtue of their humanity (Tasioulas, 26. Donnelly, 10). For that reason, HR are inherited and inalienable. In consequence, "the possession of HR cannot be conditional on some conduct or achievement of the right-holder, a relationship to which they belong, or their membership of a particular community or group" (Tasioulas, 37). HR are inherited, so they do not need to be granted by any authority, but authorities must recognize and protect them (Jones). There is no agreement on what the most basic catalogue of HR is and to what extent they may vary in different societies (Wolterstroff 2008, 314. Tasioulas, 37. Sen.) However, there is a broad consensus that there are some basic, fundamental rights common to all people (Donnelly, 94. McCrudden, 667).

HR are considered fundamental rights on which the whole legal system is based in many jurisdictions. Some HR are listed in the constitutions, and they enjoy the status of constitutional rights. Multiple detailed regulations are sourced in the fundamental rights, i.e., procedural rules of criminal proceedings stem from the right to access to the court and the right to defence (Johnson, 260). Respect for HR is considered one of the most vital requirements of democracy, but also the legitimacy of the law (Buchanan, 5).

The supremacy of HR also has its institutional aspect. National states are bound by international legal acts like the Universal Declaration of Human Rights (UDHR), the American Declaration of the Rights and Duties of Man or the European Convention on Human Rights (ECHR). Both at the UN level and regional levels there are bodies responsible for monitoring and protecting HR, to keep national states responsible for HR violations. American and European systems of HR protection give individuals the right to bring their complaints to specially designated, independent bodies. In Europe, it is the European Court of Human Rights (ECtHR), whose judgements are binding for national states.

Key Features of HR Law in Europe

There are two characteristics of the European system of HR law that should be pointed out:

1) European HR law is principle-based, leaving national states some margin of appreciation.

2) Most HR do not have an absolute character, with exceptions like the right to life.

These features are essential to understand the European system of HR law, and to conduct an intended evaluation of the Declaration.

Rights listed in the ECHR have a character of legal principles, not norms. There are two main consequences of that. First, over time, the interpretation of the right may evolve, while the wording stays the same. As a consequence, the content of the right changes as well. In 2011, in Bayatyan v. Armenia, the ECtHR recognized the right to conscientious objection for the first time, considering social changes, and “developments both in the domestic legal systems of Council of Europe member States and internationally” (para 101). Second, states must recognize all rights, but they have some margin of appreciation when it comes to the form of regulations. As McGoldrick notices, “in terms of whether the margin of appreciation applies and its width, it will be significant if the relevant law or policy is considered to reflect the ‘profound moral views of the people of the state’ (A.S. v Switzerland, para 241) or ‘concerns a question about the requirements of morals’ (Stübing v Germany, para 61)” (22). This margin of appreciation makes the law more flexible and allows it to respect social and cultural context, but also technological development (Shany. Hallström).

Most rights under the ECHR do not have an absolute character. There are only four rights that cannot be derogated: the right to life, freedom from torture or inhumane or degrading treatment or punishment, prohibition of slavery or servitude, and a prohibition of punishing without legal provision (Articles 2, 3, 4(1), 7). Legitimate limitations of rights and balancing them against each other are allowed under the ECHR. Balancing is a frequent practice at the ECtHR and requires judges to consider all interests involved (Cali). In the context of AT, it would mean that not only the interests of users and providers of AT shall be considered, but also caregivers, relatives, the general population, and the state. However, special consideration shall be given to the rights of vulnerable groups, i.e., people with disabilities (McGoldrick, 24. Alajos Kiss v Hungary, para 42).

Legal Analysis of the European Declaration on Digital Rights and Principles

A legal analysis of various aspects of the Declaration is needed to assess whether it has the potential to have a position comparable to HR law. This research will be conducted in four steps. Firstly, it is vital to acknowledge the legal position of the Declaration, and by extension of digital rights (DR). Secondly, material analysis of DR will be undertaken. Thirdly, the possible enforcement of DR will be discussed. Finally, the Declaration will be compared with the ECHR.

The Legal Status of the Declaration

The European Declaration on Digital Rights and Principles for the Digital Decade is a proposed joint document of three key institutions of the European Union: the European Parliament, the Council, and the Commission. Although the Commission has a right to legislative initiative, the Declaration does not constitute a legislative proposal *sensu stricto*. As the Declaration explains, it has a "declaratory nature and does not affect the content of legal rules or their application" (no 7 preamble), but rather is about "explaining political intentions" (no 5 Preamble). In its communication, the Commission explains that the Declaration should serve "citizens, businesses, public administrations and policy-makers" as a "guidance for a human-centred, secure, inclusive, and open digital environment" (1). The position of the Declaration should not be underestimated. There is a long-lasting tradition of interinstitutional agreements between the EU institutions, which may be of binding nature (Tournepiche. Driessen). Lenearts et al. notice that "where the Treaties do not expressly provide for an agreement to be reached, the legal force of an agreement will depend on whether the institutions intended it to be binding" (27.0500). Monar observes that if institutions do not want the agreement to be binding, they expressly state so (700-702). Therefore, although the Declaration is not binding for the EU States or individuals, it may bind institutions entering into the agreement.

Material Analysis of Digital Rights

The second step of the legal evaluation of DR is their material analysis. The main issue is whether DR are precise enough to be considered claimable rights. Secondary questions are if DR have the character of principles or norms, if they impose positive or negative obligations, and if the rights pertain to national states, the European Union, or other entities.

The declaration contains a preamble and six chapters, four of which are divided into subchapters. Each subchapter and undivided chapters consist of a statement followed by commitments. Commitments correspond with statements, and present EU policies that aim to achieve goals set out in the statements. The statements have various characters. Some of them are phrased in the language of rights: "everyone has the right…" or "nobody is to be asked to…". Other statements employ "should", and they express recommendations or the aims of the EU institutions. Out of twenty-three statements, seven are phrased in the language of rights, and sixteen in the language of recommendations. The seven rights listed in the Declaration are already guaranteed by other legal documents. The decision to phrase other statements as recommendations is in line with the intention of the Commission to prepare a declaratory document. The character of declaratory statements may be easily changed in the future by substituting "should" with "shall".

The main question remains if the statements are precise enough to allow one to formulate a claim based on them. It is important to notice that fundamental rights are rarely as precise as statutory or procedural rights. Tharney et al. notice that the law is precise if it allows "ordinary people" to understand their obligations (353). Coleman suggests that such a test is appropriate rather for clarity than for the precision of the law (407-408). There is also no clarity on what "ordinary" or "average" could mean (Cohen). The history of HR law in Europe proves that the absolute precision of the law is a myth, and only case law may bring more clarity (Mellinkoff, 424). The language of the Declaration is not far from that of the ECHR or UDHR, and most DR are phrased with a precision comparable to HR law. For that reason, it is justified to call all statements "rights" or DR.

DR are principles rather than norms, as they are phrased in a relatively general way. Some of them are vaguer than others, i.e., “Everyone should be empowered to benefit from the advantages of artificial intelligence by making their own, informed choices in the digital environment while being protected against risks and harm to one’s health, safety and fundamental rights” (Chapter III). The rights connected with privacy and individual control of data and with the protection of children are the most precise ones (Chapter V). Some DR are negative rights (freedom of expression online, confidentiality of communication), and some are positive (right to access to digital technologies, right to education, training and lifelong learning). It is worth noticing that some negative rights of citizens come along with the positive obligation of service providers or states, i.e., Chapter IV guarantees everyone the right to freedom of expression online and obliges “very large online platforms” to “support free democratic debate online”, and to “mitigate the risks stemming from the functioning and use of their services”.

The Declaration states that “the promotion and implementation of the digital principles is a shared political commitment and responsibility of the Union and its Member States”. Commitments are addressed to the EU institutions that are cosignatories of the Declaration. Most DR concern issues remaining within the Member States’ competencies. It is also clear that the implementation of DR will impose new obligations on businesses, but also individuals. Because of the status of the Declaration, it may be binding only for EU institutions. For other entities, the Declaration may be a point of reference, and an inspiration for research or industrial standards.

Enforcement of Digital Rights

The existence of effective enforcement of rights is crucial for the ability of people to enjoy them. Without that, the rights will simply remain an idealistic manifesto. (Bilchitz, 4-6). The problem of enforcement is especially vital for people from vulnerable groups, including senior citizens, and people with disabilities. Previous discussions show that currently the Declaration itself cannot be the source of legal claims. However, some DR are guaranteed through other legal acts, such as ECHR, the Charter of Fundamental Rights of the EU (Charter) or the General Data Protection Regulation (GDPR). These rights may be a source of claims enforceable in judicial proceedings, or a special procedure in the case of GDPR (Article 77, 82-84). Other DR may result in the liability of the EU institutions. Such liability would be primarily political, but because the Declaration may be legally binding, a judicial remedy cannot be ruled out.

Declaration and European Convention on Human Rights

To answer the question of whether DR may assume a role comparable to HR, it is beneficial to compare the Declaration with the ECHR. Three essential elements will be taken into consideration: legal character, precision, and enforcement mechanisms.

The Declaration and the ECHR have different legal characters and statuses. The ECHR is an international convention, legally binding for the states that ratified it. The Declaration is a declaratory document, a kind of policy statement of EU institutions, which may bind them as an interinstitutional agreement.

The precision of the language of both documents is similar. The scope and meaning of DR would have to be decided by the courts, as it has been with ECHR. The main difference in wording is that all rights from the ECHR are phrased as rights, while sixteen DR employ language more appropriate for recommendations or guidelines. It is a result of the character of the Declaration and can be smoothly amended in the future. DR are phrased precisely enough to consider them as fundamental rights.

There is an essential difference between the possible enforcement of rights from the ECHR and the Declaration. ECHR establishes the ECtHR and provides the basic rules of its proceedings. It also grants individuals a right to bring their complaints to national courts, and the ECtHR. The Declaration does not mention the issue of enforcement.

Despite the differences in legal character and provided enforcement mechanisms, there is an essential similarity between ECHR and the Declaration. They both aim to guarantee rights considered to be fundamental at the time, and in this way establish principles of the legal order.

Conclusions

Research presented in the previous parts of the paper allows the formulation of the following conclusions:

1. The Declaration is a policy statement of the EU institutions and should be read as such. It may be legally binding only for the signatories of the Declaration. However, even nonbinding documents are often used in EU policymaking as a point of reference. DR may be also used as a guideline by the Member States, industry, civic society organisations, and researchers.
2. Innovative technologies often require a regulatory response, and it may be necessary to introduce new fundamental rights in that context. Such a need and the content of these fundamental rights should be investigated in further research in the legal but also sociological and political domains. The alternative is to interpret existing HR more flexibly. However, this may not be clear to potential users of AT, who are especially vulnerable, and shall not only be protected but also feel safe.
3. If DR are to become the HR of the digital age, three crucial improvements should be made:

a. First, DR must be formulated in a precise way allowing for the formulation of a claim.

b. Second, the Declaration must be given a binding legal form, appropriate to the context and gravity of the matter. In my opinion, DR should be proclaimed at least in the form of a regulation. However, the optimal solution would be to include them in the EU primary law.

c. Thirdly, the history of HR shows how crucial a well-working enforcement mechanism is. A mechanism analogous to the one of the ECHR should be introduced, guaranteeing individuals' right to bring their complaints to the Court of Justice of the European Union, or another specialised court that would be established.

In general, the Declaration is a step in the right direction. But further steps are required to make DR more precise, binding, and enforceable. DR established in that way may resonate and cause the so-called “Brussels effect” and also impact non-Europeans, as it was with GDPR (Gunst and De Ville. Bradford). That is a possible path for DR to become the HR of the digital decade.

Acknowledgments

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 861091. This document reflects the views only of the authors, and the European Union cannot be held responsible for any use which may be made of the information contained therein.

The author thanks Ioannis Tsitsovits (KU Leuven) and Luca Love (Karolinska Institutet) for reading the previous version of this paper, and for their valuable comments.

Works Cited

Ake-Kob, Alin, et al. *State of the Art on Ethical, Legal, and Social Issues Linked to Audio- and Video- Based AAL Solutions.* COST Action GoodBrother, 2021.

Bilchitz, David. *Poverty and Fundamental Rights: The Justification and Enforcement of Socio-Economic Rights*. Oxford University Press, 2008.

Bradford, Anu. *The Brussels Effect: How the European Union Rules the World*. Oxford University Press, 2021.

Buchanan, Allen E. *Justice, Legitimacy, and Self-Determination: Moral Foundations for International Law*. Oxford University Press, 2012.

Cali, Başak. "Balancing Human Rights? Methodological Problems with Weights, Scales and Proportions." *Human Rights Quarterly*, vol. 29, no. 1, 2007, pp. 251–270., https://doi.org/10.1353/hrq.2007.0002.

Calo, Ryan. "Privacy, Vulnerability, and Affordance." *The Cambridge Handbook of Consumer Privacy*, pp. 198–206., https://doi.org/10.1017/9781316831960.011.

Cohen, Jason. "Bringing down the Average: The Case for a "less Sophisticated" Reasonableness Standard in US and EU Consumer Law." *Loyola Consumer Law Review*, vol. 32, no. 1, 2019, pp. 1-44. HeinOnline.

Coleman, Brady. "Are Clarity and Precision Compatible Aims in Legal Drafting." *Singapore Journal of Legal Studies*, vol. 1998, no. 2, 1998, p. 376-408. HeinOnline.

Council of Europe. "Convention for the Protection of Human Rights and Fundamental Freedoms." *Council of Europe Treaty Series 005*, Council of Europe, 1950. https://www.echr.coe.int/documents/convention_eng.pdf.

Custers, Bart. "New Digital Rights: Imagining Additional Fundamental Rights for the Digital Era." *Computer Law & Security Review*, vol. 44, 2022, p. 105636., https://doi.org/10.1016/j.clsr.2021.105636.

Donnelly, Jack. *Universal Human Rights in Theory and Practice*. Cornell University Press, 2013.

Driessen, Bart. *Interinstitutional Conventions in EU Law*. Cameron May, 2007.

European Court of Human Rights. *Alajos Kiss v Hungary*. 27 Mar. 2008.

European Court of Human Rights. *A.S. v Switzerland*. 30 June 2015.

European Court of Human Rights. *Bayatyan v Armenia*. 7 July 2011.

European Court of Human Rights. *Stübing v Germany*. 24 Sept. 2012.

European Union. "Charter of Fundamental Rights of the European Union. *Official Journal of the European Union C326*, European Union, 2012, pp. 391-407. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A12012P%2FTXT.

European Union, Council of the European Union and European Parliament. "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)". *Official Journal of the European Union 119,* European Union, 2016, pp. 1-88. https://eur-lex.europa.eu/eli/reg/2016/679/oj.

European Union, European Commission, "European Declaration on Digital Rights and Principles for the Digital Decade" *Shaping Europe's Digital Future*, 26 Jan. 2022, https://digital-strategy.ec.europa.eu/en/library/declaration-european-digital-rights-and-principles.

Freeman, Michael. "The Philosophical Foundations of Human Rights." *Human Rights Quarterly*, vol. 16, no. 3, 1994, p. 491., https://doi.org/10.2307/762434.

Gunst, Simon, and De Ville, Ferdi. “The Brussels Effect: How the GDPR Conquered Silicon Valley.” *European Foreign Affairs Review*, vol. 26, no. Issue 3, 2021, pp. 437–458., https://doi.org/10.54648/eerr2021036.

Hallström, Pär. “Balance or Clash of Legal Orders—Some Notes on Margin of Appreciation.” *Human Rights in Contemporary European Law*, https://doi.org/10.5040/9781474201414.ch-005.

Hayden, Patrick. *The Philosophy of Human Rights*. Paragon House, 2001.

Johnson, Brian D. “The Multilevel Context of Criminal Sentencing: Integrating Judge- and County-Level Influences*.” *Criminology*, vol. 44, no. 2, 2006, pp. 259–298., https://doi.org/10.1111/j.1745-9125.2006.00049.x.

Jones, Peter. “The Philosophical Foundation of Human Rights.” *National Rights, International Obligations*, edited by Simon Caney et al., Westview Press, Boulder, CO, 1996.

Lenaerts, Koen, et al. *EU Constitutional Law*. Oxford Academic, 2022.

Limantė, Agnė. “Vulnerable Groups in the Case Law of the European Court of Human Rights.” *Legal Protection of Vulnerable Groups in Lithuania, Latvia, Estonia and Poland*, edited by Agnė Limantė and Dovilė Pūraitė-Andrikienė, Springer, 2022, pp. 29–54.

Mathiesen, Kay. “Human Rights for the Digital Age.” *Journal of Mass Media Ethics*, vol. 29, no. 1, 2014, pp. 2–18., https://doi.org/10.1080/08900523.2014.863124.

McCrudden, C. “Human Dignity and Judicial Interpretation of Human Rights.” *European Journal of International Law*, vol. 19, no. 4, 2008, pp. 655–724., https://doi.org/10.1093/ejil/chn043.

McGoldrick, Dominic. “A Defence of the Margin of Appreciation and an Argument for its Application by the Human Rights Committee.” International and Comparative Law Quarterly, vol. 65, no. 1, 2016, pp. 21–60., doi:10.1017/S0020589315000457.

Mellinkoff, David. "The Myth of Precision and the Law Dictionary." *UCLA Law Review*, vol. 31, no. 2, December 1983, pp. 423-442. HeinOnline.

Monar, Jörg. “Interinstitutional Agreements: The Phenomenon and Its New Dynamics after Maastricht.” *Common Market Law Review*, vol. 31, no. Issue 4, 1994, pp. 693–719., doi:10.54648/cola1994037.

Sen, Amartya. *The Idea of Justice*. The Belknap Press of Harvard University Press, 2009.

Shany, Yuval. “All Roads Lead to Strasbourg?: Application of the Margin of Appreciation Doctrine by the European Court of Human Rights and the UN Human Rights Committee.” *Journal of International Dispute Settlement*, vol. 9, no. 2, 2017, pp. 180–198., https://doi.org/10.1093/jnlids/idx011.

Solove, Daniel J. *Understanding Privacy*. Harvard University Press, 2010.

Tasioulas, John. “On the Nature of Human Rights.” *The Philosophy of Human Rights: Contemporary Controversies*, by Gerhard Ernst and Jan-Christoph Heilinger, De Gruyter, 2012, pp. 17–59.

Tharney, Laura C., et al. "On the Path toward Precision: Responding to the Need for Clear Statutes in the Criminal Law." *Seton Hall Legislative Journal*, vol. 45, no. 2, 2021, pp. 329-354. HeinOnline.

Tournepiche, Anne-Marie. *Les Accords Interinstitutionnels Dans L'Union Européenne*. Bruylant, 2011.

United Nations, General Assembly. “Universal Declaration of Human Rights”. 1948. https://www.un.org/en/about-us/universal-declaration-of-human-rights.

Wolterstorff, Nicholas. *Justice: Rights and Wrongs*. Princeton University Press, 2010.

An AI-enabled Annotation Platform for Storefront Accessibility and Localization

Xuan Wang[3], Jiawei Liu[1], Hao Tang[2,3], Zhigang Zhu[1,3], William H. Seiple[4]
[1]Department of Computer Science, The CUNY City College
[2]Department of Computer Information Systems, The CUNY BMCC
[3]Department of Computer Science, The CUNY Graduate Center
[4]Lighthouse Guild, New York, NY 10023
xwang4@gradcenter.cuny.edu, jliu021@citymail.cuny.edu,
htang@bmcc.cuny.edu, zzhu@ccny.cuny.edu, wseiple@lighthouseguild.org

Abstract

Although various navigation apps are available, people who are blind or have low vision (PVIB) still face challenges to locate store entrances due to missing geospatial information in existing map services. Previously, we have developed a crowdsourcing platform to collect storefront accessibility and localization data to address the above challenges. In this paper, we have significantly improved the efficiency of data collection and user engagement in our new AI-enabled *Smart DoorFront* platform by designing and developing multiple important features, including a gamified credit ranking system, a volunteer contribution estimator, an AI-based pre-labeling function, and an image gallery feature. For achieving these, we integrate a specially designed deep learning model called *MultiCLU* into the Smart DoorFront. We also introduce an online machine learning mechanism to iteratively train the MultiCLU model, by using newly labeled storefront accessibility objects and their locations in images. Our new DoorFront platform not only significantly improves the efficiency of storefront accessibility data collection, but optimizes user experience. We have conducted interviews with six adults who are blind to better understand their daily travel challenges and their feedback indicated that the storefront accessibility data collected via the DoorFront platform would be very beneficial for them.

Keywords

Crowdsourcing, Storefront Accessibility, Independent Travel, Visually Impaired, Open-Source Data, Deep Learning.

Introduction

As reported by multiple sources (CDC, 2021; IAPB, 2022), over four million people living in the United States are blind or have low vision, and the prevalence of vision loss is predicted to increase dramatically in the next decades. The limitations caused by vision loss in daily life are unimaginable, even travel can be considered as the one of the most stressful events in the lives of people with visual impairment or blindness (PVIB) (Donaldson, 2017). Difficulties with travel limit essential activities, such as visiting local stores and using public transportation facilities, etc. In addition to the obstacles presented by many transportation options, proximal information about storefronts is essential for access and poorly designed entrances place PVIB at risk (MOPD, 2022; ADA, 2022). Therefore, knowing the accessibility data of the storefronts in advance can significantly alleviate their fear in exploring their local community.

In order to gather storefront accessibility and localization image data in a complicated street level environment, to support independent travel of PVIB, we have developed *DoorFront* (Liu, et al., 2022), a crowdsourcing web platform that collects large-scale storefront accessibility and localization data in New York City. Our web-based platform DoorFront allows volunteers to remotely label storefront accessibility objects on Google Street View panoramas (Google, 2022) in four main categories: *door*, *doorknob, stair* and *ramps*. The feedback from volunteers have demonstrated its usability and high potential for data collection.

However, the data collection of large-scale storefront accessibility and localization data using crowdsourcing is still very labor intensive. Therefore, in this paper, we introduce an AI-enabled annotation platform for storefront accessibility and localization, which is built on our previous DoorFront platform. In order to improve user experience, user label efficiency and quality, we implemented a number of new features into the DoorFront platform, including gamified credit ranking, volunteer contribution estimation, AI-based pre-labeling and image gallery. As a key component, we integrate a specially designed deep learning model called *MultiCLU* (Wang, et al., 2022), for enabling label automation. Furthermore, we introduced an iterative training mechanism for our MultiCLU model to provide better labeling performance and 'autowalker,' an enhanced pre-label process to further improve the efficiency of human labeling. To summarize, the contributions in this work include:

1. An AI-enabled annotation platform for storefront accessibility and localization, *Smart DoorFront*, which integrates new features, including credit ranking, volunteer contribution estimating, AI-based pre-labeling, and image gallery.
2. Enabling label automation by integrating a specifically designed deep learning model - MultiCLU.
3. An iterative online training mechanism for our deep learning model to provide better labeling performance and user experience.

Related Work

Crowdsourcing Accessibility Data

Accessibility data, including sidewalk accessibility data, storefront accessibility data, and public infrastructure data, is essential information for PVIB in order to plan independent travel. To collect up-to-date accessibility data, local and state governments, and even federal government, often conduct street audits to inquire about specific conditions (May et al., 2014; Law et al., 2018). However, manual audits are both time consuming and expensive. One of the novel alternatives is the use of crowdsourcing techniques. Many recent studies have demonstrated the feasibility of crowdsourcing approaches (Krajzewicz et al., 2010; Marzano et al., 2019), in collecting useful information on urban mobility and public transportation. Hara and his team have proposed methods combining Google Street View and online crowdsourcing to provide sidewalk accessibility data and bus stop locations (Hara et al., 2013, 2014). However, there are few efforts focused on storefront accessibility and localization information.

Object Detection in Urban Environment

Many approaches have been proposed for object detection in urban scenes. Some of them (Du et al. 2012; Sabir et al. 2018; Zhu et al. 2016) are for text and signage detection in street level environments. Ahmetovic, et al. (2015) mined existing spatial image datasets for discovery of zebra crosswalks in urban environments, which could increase safety of PVIB when traveling. Sun et al. (2017) used deep learning to find missing curb ramps at city street regions, which could help not only PVIB, but also people with mobility disabilities. Weld, et al. (2019) propose a deep learning framework to automatically detect sidewalk accessibility using streetscape imagery. To our best knowledge, there are few, if any, studies related to detect storefront accessibility and localize them using street level imagery. In this paper, we integrate our specifically designed deep learning model, MultiCLU (Wang et al. 2022), to enable pre-label automation for our DoorFront platform by using our previously collected storefront accessibility and localization image dataset.

Discussion

User Study

In order to ensure the effectiveness and usefulness of our Smart DoorFront platform, we conducted informal interviews with six adults who are blind to better understand their challenges when they accessed essential activities. Table 1 shows the doorfront interview questions and the answers from the interviewees. We started with basic travel questions (Q1 to Q4 in Table 1), followed by storefront accessibility related questions (Q5 to Q11 in Table 1).

Table 1. Doorfront Interview Questions and Answers.

Questions	**Answers**
1.How often do you travel?	Every day: 100% Once a week: 0% Once a month: 0%
2.What is your major transportation while you are traveling?	Access-a-Ride: 17% Subway: 83% Walking: 0%
3.How far do you often travel?	<1 mile: 0% 1 - 5 miles: 83% > 5 miles: 17%
4. How long do you spend per trip?	< 1 hour: 17% 1 - 3 hours: 83% 3 - 5 hours: 0% > 5 hours: 0%
6.Which store do you often visit (Multiple choices)?	Grocery store: 100% Supermarket: 67% Restaurant: 0% Shopping mall: 0%
7.How difficult is it to find the entrance of the store (Rate from 1 to 5)?	1 (very easy): 0% 2 (easy): 0% 3 (normal): 0% 4 (hard): 83% 5 (very hard): 17%
8.What is the most challenging task when you arrive at the entrance of a store?	Locate the entrance; Door type for the entrance. Time consuming if no one helps.
9.What kind of assistance do you expect when you arrive at the entrance of a store?	Direction guidance: 17% Lead to the entrance: 83%
10.Our current application can locate and detect three main storefront objects: door, knob and stair, with estimated location of knob, do you think they are sufficient?	Yes: 100% No: 0%
11.What do you expect to know when you arrive at the entrance of a store?	Accurate location for the door and stair. Relative location for the knob.

Through this series of interviews, we learned that most of the participants were daily commuters, and that they travel independently. All participants often visited grocery stores or supermarkets, but none of them visited restaurants and shopping malls (Q6 in Table 1). They stated that finding the locations of store entrances is very challenging for them (Q7 in Table 1). Two participants mentioned their previous experiences of finding a store entrance. They specifically indicated that one of the most challenging tasks is to find the entrance with a glass door. There are many buildings with glass walls at the ground floor in NYC, our participants indicated that it is very challenging to find the glass door that hides inside the glass walls. If no one provides help, they could spend a great amount of time finding and locating the entrance. During the interview, we found that each participant needs different levels of assistance to find the entrance of a store, but most of them prefer a person who can lead them to the entrance directly, instead of providing verbal guidance or direction. Most participants indicated that knowing the location of the entrance (e.g., a door) and the access of the entrance (e.g., the stairs to the door) were very important for them, which could significantly reduce their travel time and relieve their stress. They also agreed that providing a relative location for the doorknob to the door (e.g., "*The knob is on the left side of the door*.") could work for them if the location for the entrance is accurate.

The feedback from the interviews indicated that helping PVIB to identify the accessibility of the storefront would be very beneficial for them, and could greatly ease their daily burdens and improve their independence. Not only we received valuable feedback for better understanding challenges for PVIB through the survey, but we also obtained ideas on expanding the categories of storefront accessibility objects (e.g., handrails for the stairs) and improving our platform.

Doorfront Platform Improvements: An Overview

DoorFront is a web-based application that combines Google Street View and crowdsourcing, with an interactive interface and a user-friendly labeling tool (Liu, et al., 2022). There are two main pages in the original DoorFront, namely an *Exploration* page and a *Labeling* page. Volunteers not only can virtually walk-through New York City with embedded interactive Google Street View provided on the Exploration page; but they can also label storefront accessibility data with the functional and user-friendly labeling tool. Even though the feedback from crowd volunteers has demonstrated its usability and high potential for data collection, the process of labeling is still relatively labor intensive. Our studies show that there are two key factors that influence the effectiveness of the data collection: the number of volunteers participating in DooFront and the time they spent in the labeling process.

To address these issues, we have made significant improvements to the DoorFront application, leading to *Smart DoorFront,* which includes four major new features: gamified credit ranking, volunteer contribution estimation, AI-based pre-labeling and image gallery. We will discuss each of them in the following.

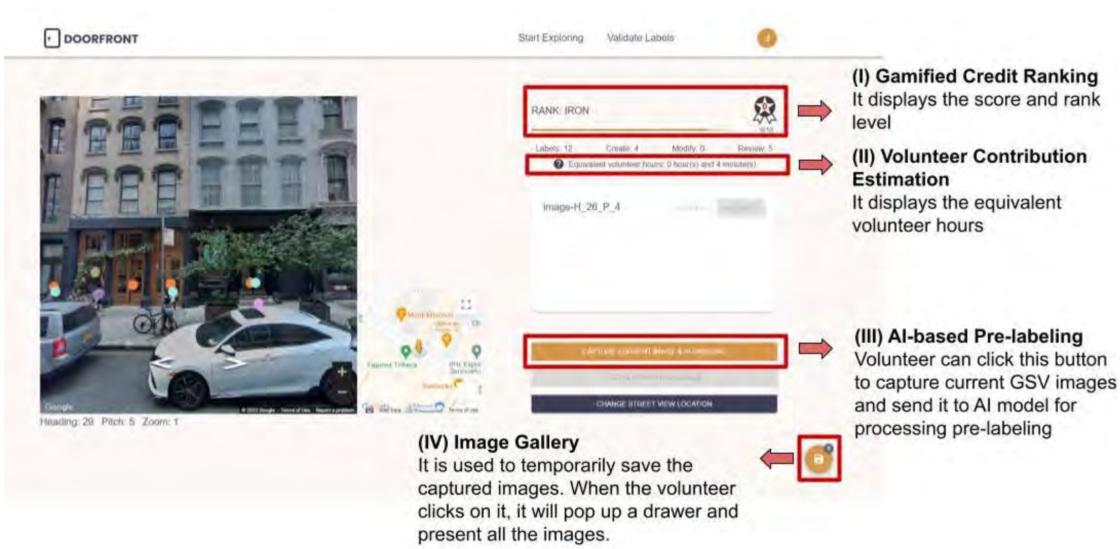


Fig. 1. The Interface of the *Smart DoorFront* Exploration Page.

Gamified Credit Ranking

Inspired by gamified settings (Ning, 2018), we designed and implemented a *gamified ranking system* (Fig 1, part I) and a leaderboard for the volunteers. In this project, we define seven different rank levels and their corresponding cumulative credits. The specific names of each level and the details of the accumulated credits for each phase are shown in Table 2.

Table 2. Rank levels and accumulated credits.

Level	**Iron**	**Bronze**	**Silver**	**Gold**	**Platinum**	**Diamond**	**Challenger**
Credits	0 - 9	10 - 49	50 - 299	300 - 999	1000 - 1999	2000 - 4999	5000 - 9999

Volunteers will receive credits through three contributed operations: annotating new Google Street View images, correcting other volunteers' annotations, and reviewing other volunteers' annotations. To further increase the entertaining nature of DoorFront, we also develop a treasure hunt feature that allows volunteers to earn extra ten credits whenever they find a treasure; once DoorFront is initialized, all treasures will automatically be hidden in random areas of New York City. Therefore, the locations of the treasures are completely different each time, and volunteers are unable to gain these extra points by memorizing the locations of the treasures. With respect to leaderboard, it will show volunteers according to their levels. In this atmosphere, we believe volunteers will become competitive and spend more time in data collection to advance their level.

Volunteer Contribution Estimation

Crowdsourcing brings us flexibility so we can distribute the data collection tasks to volunteers. Volunteers from each borough in New York City can collect storefront accessibility data in their own communities through DoorFront. However, using the original DoorFront platform, we were unable to recruit a large number of volunteers to participate. We needed to address how to attract more volunteers to DoorFront. In the initial version, we decided to award volunteer certificates to volunteers through our collaboration with Lighthouse Guild, a vision and health care organization. They provided volunteer certificates to individuals who have made large contributions of time on the platform. However, the algorithm we used did not work well to calculate the equivalent volunteer time. The core idea was calculating the number of images collected by the volunteers. We assume that, on average, volunteers spend one minute to annotate each image without the assistance of the AI model. The shortcoming of this algorithm is obvious. Since the number of labels in each image is different, our algorithm does not reflect well the effort of volunteers and the time they spend.

Therefore, we decided to design a new *volunteer contribution estimator* to better calculate volunteer effort. First, we rebuilt the DoorFront's credit management system. With this improvement, we are now able to monitor the number of labels annotated by volunteers, and equivalent-volunteer-time is determined by the number of labels. Second, we implemented a small widget to showcase the volunteer’s effort in real-time (Fig. 1, part II). In addition, to further encourage volunteers to promote our application, we also provide sharing buttons on different social media applications such as Meta and Twitter, to share their contributions with their friends. With these improvements, we believe that more and more younger volunteers, especially middle or high school students, will be interested in participating in our study.

AI-based Pre-labeling

One of the key issues we needed to address is to reduce the time for annotation by a volunteer. On average, it takes at least one minute to manually annotate a storefront image from scratch using our DoorFront interface. There are three steps in the annotation process: (1) identify a storefront accessibility object; (2) annotate the object with a bounding box; and (3) add a subtype if the object is a door or door handle. Volunteers need to repeat these three steps until they label all the storefront accessibility objects in a scene, hence the labeling task is still time-consuming.

In order to further improve the efficiency of data collection, we enhanced *DoorFront* with an *AI-based pre-labeling* function (Fig. 1, part III). With AI support, DoorFront can perform pre-labeling once a volunteer captures a new Google Street View image. This means that they do not need to label the image from scratch and the only thing they need to do is validate the results predicted by our AI model. Compared to the initial workflow, we now have only two steps: (1) Verify and correct the annotations labeled by the AI model; and (2) Add subtypes. Based on the outstanding performance of our model, we can skip the first step in most cases, which dramatically reduces the annotation time.

In addition, we introduced an *autowalker*, a pre-label process to further improve the efficiency of human labeling. Our autowalker will automatically walk through the Google Street View in the direction selected by a volunteer, capture 50 images, and use our model to pre-label the storefront accessibility data. Meanwhile, our autowalker will also collect the GIS information (latitude, longitude) and corresponding heading, pitch and zoom level.

Image Gallery

To maximize the utilization of the AI model, we modified the way that we save Google Street View images. In the initial version, DoorFront only allowed volunteers to label one image at a time. Now with the Smart DoorFront, volunteers can capture multiple images while they are virtually walking along with the street. Those images will be temporarily stored in an *image gallery* (Fig 1, part IV) and then sent to the AI model for pre-labeling processing. Volunteers can then validate all images at once, without frequently switching among different web interfaces, which greatly reduces their labeling time.

Furthermore, we store the untagged images in our remote database. With this information, volunteers will be able to access these images again, regardless of the last time they exited the application. Furthermore, our application will send notifications to remind volunteers that they forgot to annotate these images.

In the next sections, we will describe the enabler of the aforementioned features: the integration of the deep learning model and an iterative learning approach.

Deep Learning Model Integration

In order to improve the efficiency of the storefront accessibility labeling process, we integrate a specially designed deep learning model MultiCLU (Wang, et al., 2022) into DoorFront. Our MultiCLU model is implemented by integrating the state-of-the-art object detector, Faster R-CNN (Ren, et al., 2015), with contextual relationships among storefront accessibility objects, in order to improve the accuracy of image detections (Fig. 1, part III). For example, if we know there is a knob in the image, we can easily guess there should be a door in the image. Furthermore, a knob must appear inside a door frame, and if a stair exists, it should be under the door. Our MultiCLU model utilizes these contextual relationships to improve our detection results. The overall pipeline of MultiCLU in DoorFront is shown in Fig. 2. When a volunteer captures an image, MultiCLU will detect and localize storefront accessibility objects within a few seconds before user labeling. Volunteers can then validate or edit the labels which are pre-labeled by MultiCLU. Our platform will record three main types of labels: (1) Added labels from volunteers (Fig. 2. part I); (2) Removed labels from volunteers (Fig. 2. part II); and

(3) Validated AI labels (Fig. 2. part III). Our model is further trained on modified labels which are corrected by volunteers, to further improve the performance.

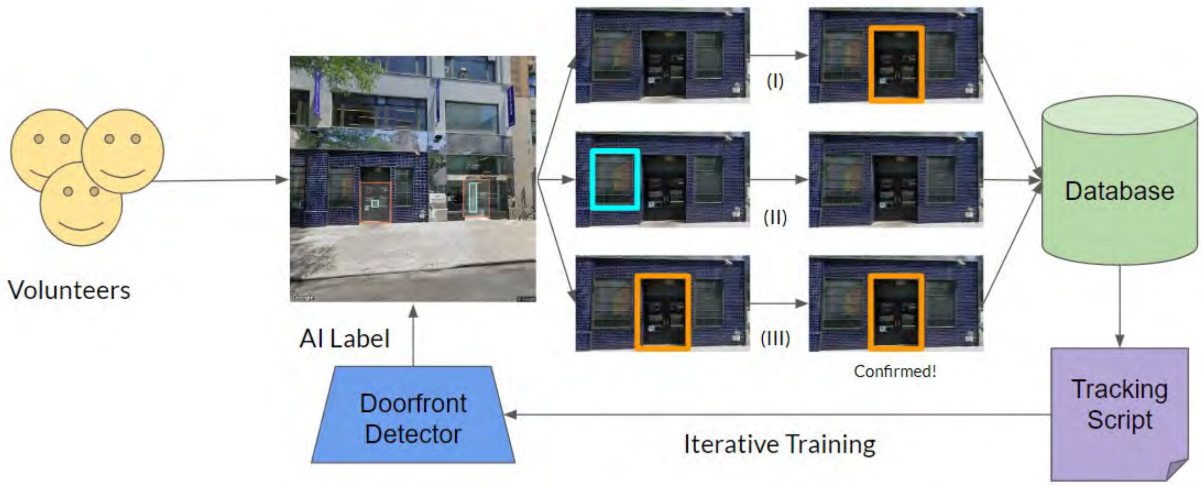


Fig. 2. Overall pipeline of MultiCLU-based labeling in DoorFront.

Dataset Description

We used 1232 storefront accessibility images previously collected by our DoorFront platform for training our initial MultiCLU model. Among them, there are 604 images which have labels from multiple volunteers. We used the average labels from these images as our base training set. The remaining images were used as a testing set. We did not use the ramp category for our training because of the lack of the labels. Table 3 and Table 4 show the summary of the dataset. The precision and recall for the three categories are shown in the “Initial Model” in Table 5 and Table 6, respectively.

Table 3. The training set of collected storefront accessibility dataset over time.

Class	Initial Training set	After Day 1	After Day 2	After Day 3
Door	1225	1532	1719	1913
Knob	863	962	1060	1173
Stair	270	346	422	475

Table 4. The testing set of collected storefront accessibility dataset over time.

Class	**Initial Testing set**	**After Day 1**	**After Day 2**	**After Day 3**
Door	1080	1132	1168	1229
Knob	887	905	928	979
Stair	197	209	224	240

Iterative Training

We also introduced a training automation mechanism to iteratively train our MultiCLU model. The model will start iterative training automatically when a certain amount (***N***) of new images has been recorded with new labels. As shown in Fig. 3, if we achieve better performance after training, we will replace the current model, otherwise we will start another training process the next day and keep the current model. We also use a data aggregation process to improve the robustness of the detection. When the $(n+1)$/ [/?] *th* iteration starts, we accumulate the previous training dataset from *n* [/?] *th* training step, where *n* denotes current training step.

And then we use the combined dataset to refine the current model.

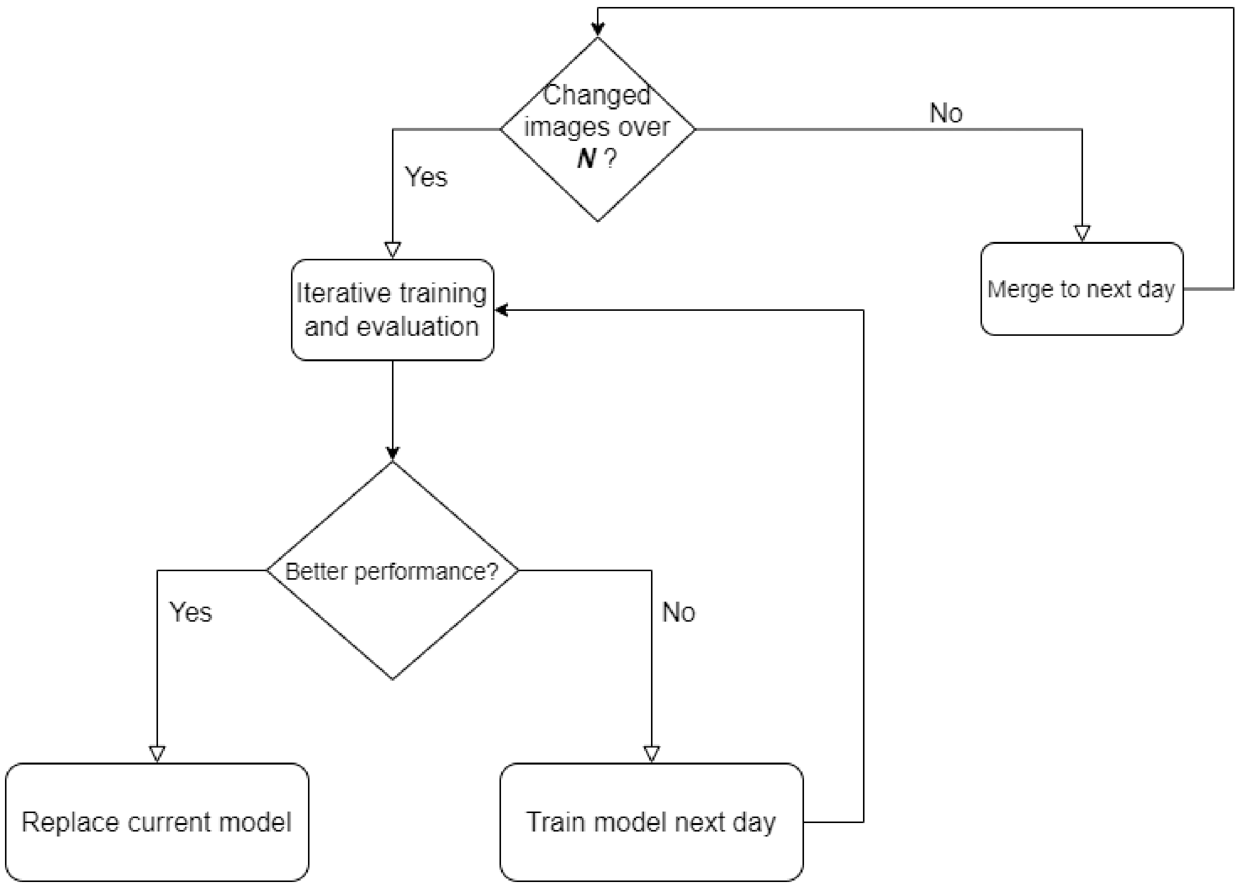


Fig. 3. Iterative training process with MultiCLU and DoorFront.

Model Evaluation

We evaluated our iterative training mechanism using labels collected within three consecutive days. For new labels from each day, we randomly select 80 percent of the dataset as the training set to refine the previous MultiCLU model, and the remaining 20 percent were added to the current testing set (Table 3 and Table 4). We accumulated both training and testing data into our previously collected storefront accessibility and localization dataset. We only used the accumulated labels from three consecutive days to refine our model, which could improve the robustness of the detection. We report both precision (Table 5) and recall (Table 6) of the 3-day iterative training. We observed that both precision and recall for all the categories were improved, with precision increasing from +1.1% to +3.7% and recall from +1.3% to +5.2%, respectively.

Table 5. The precision for the initial model and 3-day iterative trained models.

Class	Initial Model	Day 1	Day 2	Day 3
Door	78.7%	79.5%	83.2%	83.2%
Knob	79.0%	78.8%	81.6%	82.7%
Stair	81.2%	81.2%	81.6%	82.3%

Table 6. The recall for the initial model and 3-day iterative trained models.

Category	Initial Model	Day 1	Day 2	Day 3
Door	88.2%	88.9%	89.7%	90.2%
Knob	85.4%	85.4%	86.7%	88.0%
Stair	77.6%	78.0%	82.6%	83.2%

Conclusions

Building on our previous DoorFront platform, our new platform, Smart DoorFront, integrates gamified credit ranking, volunteer contribution estimation, AI-based pre-labeling and image gallery. We utilize our specially designed storefront image detection model MultiCLU, which is built upon the state-of-the-art object detector and uses of context information among storefront accessibility objects. In addition, we introduce an autowalker, a pre-label process to further improve the efficiency of human labeling. We also introduce an iterative training mechanism to improve the accuracy and robustness of our deep learning model. Our new platform not only optimizes user experience, but also significantly improves the efficiency of storefront accessibility labeling process with our deep learning model. We will continue gathering feedback from volunteers and develop a mobile app for PVIB to navigate to store entrances, using the collected storefront accessibility and localization data. We will also integrate our deep learning model into a mobile app in the future, to better help PVIB to improve their independence. We will then perform a formal user evaluation of storefront accessibility and localization using the mobile app to understand how much the app could improve their performance, and what improvements we need for the algorithms and mobile app.

Acknowledgments

The research is supported by the National Science Foundation (NSF) through Awards #2131186 (CISE-MSI), #1827505 (PFI), and #1737533 (S&CC). The work is also supported by the US Air Force Office of Scientific Research (AFOSR) via Award #FA9550-21-1-0082 and the ODNI Intelligence Community Center for Academic Excellence (IC CAE) at Rutgers University (#HHM402-19-1-0003 and #HHM402- 18-1-0007).

Works Cited

Ahmetovic, Dragan, Roberto Manduchi, James M. Coughlan, and Sergio Mascetti. “Zebra crossing spotter: Automatic population of spatial databases for increased safety of blind travelers.” *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*. 2015.

The Americans with Disabilities Act Checklist for Readily Achievable Barrier Removal (ADA), https://www.ada.gov/racheck.pdf, last visited Aug 2022.

Cheng, Mingmei, Yigong Zhang, Yingna Su, Jose M. Alvarez, and Hui Kong. “Curb detection for road and sidewalk detection.” *IEEE Transactions on Vehicular Technology* 67, no. 11 (2018): 10330-10342.

Donaldson, Nadine. “Visually Impaired Individuals' Perspectives on Obtaining and Maintaining Employment.” (2017).

Du, Yuning, Genquan Duan, and Haizhou Ai. “Context-based text detection in natural scenes.” *19th IEEE International Conference on Image Processing. IEEE*, 2012.

“Fast Facts of Common Eye Disorders.” Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 9 June 2020, www.cdc.gov/visionhealth/basics/ced/fastfacts.htm

“Global Estimates of Vision Loss.” The International Agency for the Prevention of Blindness, 31 May 2022, https://www.iapb.org/learn/vision-atlas/magnitude-and-projections/global/.

Google Map API (2022), https://developers.google.com/maps, last visited Aug 2022

Hara, Kotaro. “Scalable methods to collect and visualize sidewalk accessibility data for people with mobility impairments.” *Proceedings of the adjunct publication of the 27th annual ACM symposium on User interface software and technology*. 2014.

Hara, Kotaro, Shiri Azenkot, Megan Campbell, Cynthia L. Bennett, Vicki Le, Sean Pannella, Robert Moore, Kelly Minckler, Rochelle H. Ng, and Jon E. Froehlich. "Improving Public Transit Accessibility for Blind Riders by Crowdsourcing Bus Stop Landmark Locations with Google Street View." *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, 2013, https://doi.org/10.1145/2513383.2513448.

Krajzewicz, Daniel. "Traffic Simulation with SUMO - Simulation of Urban Mobility." *Fundamentals of Traffic Simulation*, 2010, pp. 269–293., https://doi.org/10.1007/978-1-4419-6142-6_7.

Law, Stephen, Brooks Paige, and Chris Russell. "Take a Look Around: Using Street View and Satellite Images to Estimate House Prices." *ACM Transactions on Intelligent Systems and Technology*, vol. 10, no. 5, 2019, pp. 1–19., https://doi.org/10.1145/3342240.

Liu, Jiawei, Hao Tang, William Seiple, and Zhigang Zhu. "Annotating Storefront Accessibility Data Using Crowdsourcing." *Journal on Technology and Persons with Disabilities* Volume 10, California State University, Northridge., 1 Jan. 1970, https://scholarworks.csun.edu/handle/10211.3/223472.

Marzano, Gilberto, Joanna Lizut, and Luis Ocha Siguencia. "Crowdsourcing Solutions for Supporting Urban Mobility." *Procedia Computer Science*, vol. 149, 2019, pp. 542–547., https://doi.org/10.1016/j.procs.2019.01.174.

May, Andrew, Christopher J. Parker, Neil Taylor, and Tracy Ross. "Evaluating a Concept Design of a Crowd-Sourced 'Mashup' Providing Ease-of-Access Information for People with Limited Mobility." Transportation Research Part C: Emerging Technologies, vol. 49, 2014, pp. 103–113., doi:10.1016/j.trc.2014.10.007.

Ning, Bing. “Aux-driven design method for building gamification system.” International Conference of Design, User Experience, and Usability. Springer, Cham, 2018.

NYC Mayor's Office for People with Disabilities (MOPD) https://www1.nyc.gov/site/mopd/index.page, last visited Aug 2022

Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. “Faster r-cnn: Towards real-time object detection with region proposal networks.” Advances in neural information processing systems 28 (2015).

Sabir, Ahmed, Francesc Moreno-Noguer, and Lluís Padró. “Enhancing text spotting with a language model and visual context information.” *In Proceedings of the Computer and Communications Industry Association*, 2018.

Sun, Jin, and David W. Jacobs. “Seeing what is not there: Learning context to determine where objects are missing.” *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5716-5724. 2017.

Wang, Xuan, Jiajun Chen, Hao Tang, and Zhigang Zhu. “MultiCLU: Multi-stage Context Learning and Utilization for Storefront Accessibility Detection and Evaluation.” 2022 International Conference on Multimedia Retrieval, https://dl.acm.org/doi/pdf/10.1145/3512527.3531361

Weld, Galen, Esther Jang, Anthony Li, Aileen Zeng, Kurtis Heimerl, and Jon E. Froehlich. “Deep learning for automatically detecting sidewalk accessibility problems using streetscape imagery.” *In The 21st International ACM SIGACCESS Conference on Computers and Accessibility*, pp. 196-209. 2019.

Zhu, Anna, Renwu Gao, and Seiichi Uchida. “Could scene context be beneficial for scene text detection?” *Pattern Recognition* 58 (2016): 204-215.

Onomatopoeia System for Animal Sounds Using Preconceived Expressions

Chiemi Watanabe, Miki Namatame, Rumi Hiraga
Tsukuba University of Technology
chiemi@a.tsukuba-tech.ac.jp, miki@a.tsukuba-tech.ac.jp,
rhiraga@a.tsukuba-tech.ac.jp

Abstract

A system of animal sound onomatopoeia is presented. The system enables users to view animal images linked to creative onomatopoeia. Additionally, an evaluation of the prototype system by deaf or hard of hearing (DHH) university students is presented. Through the presentation of our onomatopoeia application, we would like to discuss onomatopoeia and hearing impairment while learning about the global perspective. The first objective of the application is to let DHH children understand that each person hears differently. The second objective is to get the children interested in sounds and animals. In the future, we would like to use our application as a tool to train hearing abilities and at zoos for learning support. We welcome suggestions and opinions on possible improvements to help us achieve this goal.

Keywords

Deaf or Hard of Hearing, Zoo, onomatopoeia, animal sounds.

Introduction

Conventional methods of communicating animal sounds describe the sound source, such as “the sound of a dog”, or use generalized expressions (symbolic onomatopoeia), such as “wan-wan” or “woof-woof”. This study communicates animal sounds to deaf or hard of hearing (DHH) individuals using “creative onomatopoeia”. This novel approach uses onomatopoeic words to convey sounds as they are heard and received by the listener. This approach responds to a DHH person’s desire to understand how they hear. We want to convey to DHH children with residual hearing that they should enjoy the sound, that it is okay if they do not hear a dog's sound as “woof-woof”, and that each person hears and receives sound differently. By communicating this, this study aims to help DHH children with residual hearing to become interested in sound while helping them accept their own way of hearing.

The hearing ability of DHH children can be developed more effectively with early intervention (Ramkalawan and Davis 103). The purpose of this study is to develop a system that provides not only animal sounds but also creative onomatopoeia that expresses the sounds visually, which allows DHH children with residual hearing to learn while having fun. The system will benefit scientific and developmental education as it will motivate children to learn more about animals and train their auditory skills.

Onomatopoeia is common in Japan and picture books with onomatopoeia are often read to children with hearing disabilities. In the education of DHH children in Japan, it is recommended to verbalize actions, feelings, and sounds repeatedly. The Japanese educational guidelines for DHH children introduce an episode in which a child learned about "sound" by watching onomatopoeia depicted in an animated cartoon (MEXT, 192).

The educational role of zoos in conserving biodiversity is increasing every year. In particular, exhibits that recreate the way animals live allow visitors to learn about animal sociality. It is expected that linking animal sounds with onomatopoeia will enrich the zoo experience for people at DHH. Furthermore, by linking the experience of listening to animal sounds with onomatopoeia and knowledge, it is expected to improve language expression and vocabulary skills. It is hoped that DHH children will become interested in animals and actively visit the zoo.

Methods of informing people about sound can be classified into three categories: (1) informing people of its presence, (2) informing people of what the sound is, and (3) informing people of the atmosphere created by the sound. Examples of informing people of the presence of a sound include a hairpin-shaped device, which indicates the presence of a sound by vibration or light. Furthermore, devices worn on the body allow the user to enjoy music through vibrations and sound or alert the user using vibration and light. Ontenna was developed specifically as a device for the use by DHH people (Ontenna), whereas the LIVEJACKET allows the wearer to register rhythm using the entire body (Hashizume et al. 2). Furthermore, smartphone applications can inform users about environmental sounds, and balloons have been proposed as a way to convey sound ambiance (Mylar Ballon; SoundHug). Finally, films that make music visually enjoyable through subjective media conversion, such as Fantasia, allow DHH people to imagine sounds. Methods of indicating sounds are being researched extensively and researchers involved in environmental sound classification have had to compete for accuracy using a specific data set (DCASE community).

In Japan, an experimental method has been initiated that determines the sound source and converts it into a generalized linguistic “symbolic onomatopoeia”. For example, sound cues are generated at train stations (Ekimatope). The “creative onomatopoeia” introduced in this study is an attempt to describe sounds by adding images. Therefore, this study is a combination of 2 and 3, as well as the verbalization of sound textures.

Methodology

Collection of Creative Onomatopoeia

In this study, instead of using symbolic onomatopoeia associated with animal names to describe animal sounds, we asked participants to listen to actual animal sounds and describe the sounds they heard as creative onomatopoeia.

The animal sounds were presented in the form of 30 s videos of animals making the sounds. Preschoolers were selected to create the onomatopoeic symbolization from the animal sounds because they had not yet learnt commonly used animal onomatopoeia. The preschoolers, with their parents, were asked to watch the videos obtained from the zoo and imitate the sounds as they heard them. The parents were asked to verbalize them and confirm they were correct before submission. Table 1 lists the creative onomatopoeia presented by the children. Note that the onomatopoeias were originally presented in Japanese and have been translated phonetically.

Table 1. Creative onomatopoeia.

Animal	**Onomatopoeia**
Small-clawed Otter (Lutra lutra)	kiii kukyii, pyikyuu, pīpī, pippi, miii, myāmyā, kyūkyū, uueeuuee, dī, pyēē, uuīiuuīi
Elephant (Elephantidae)	pawoon, muumuu, fuoon, fueen, eeen, fuu, piewoo,
Elephant (Elephantidae)	pawoon, muumuu, fuoon, fueen, eeen, fuu, piewoo,

Presentation of Creative Onomatopoeia

The purpose of presenting creative onomatopoeia is to allow users to imagine animal sounds while having fun. Therefore, the onomatopoeia were superimposed over the animal sound videos used. The system was written and programmed in JavaScript-based p5.js. The system retrieves the creative onomatopoeia in real-time from a spreadsheet and superimposes them onto animal videos uploaded to a video posting site. The program allows the user to freely change the font and control the speed at which the font flows. It is also possible to treat the text as an image. The system presents creative onomatopoeia as dynamic information, which appears, flows, and disappears. However, the timing and position of the creative onomatopoeia are not calibrated owing to the variety of onomatopoeia. In the future, the text should be displayed near the animal in accordance with the timing of the animal's cry. Figure 1 is a snapshot taken during the video. In this video the elephant is making noise throughout the duration of the video and the creative onomatopoeia of the cry is displayed in five ways (Pawoon, Piewoo, Fuu, Fuooon, Fueeen).

Figure 2 shows the diagram of the system concept for animal sound onomatopoeia. The system let users look at creative onomatopoeia overlapped with movies of animals.

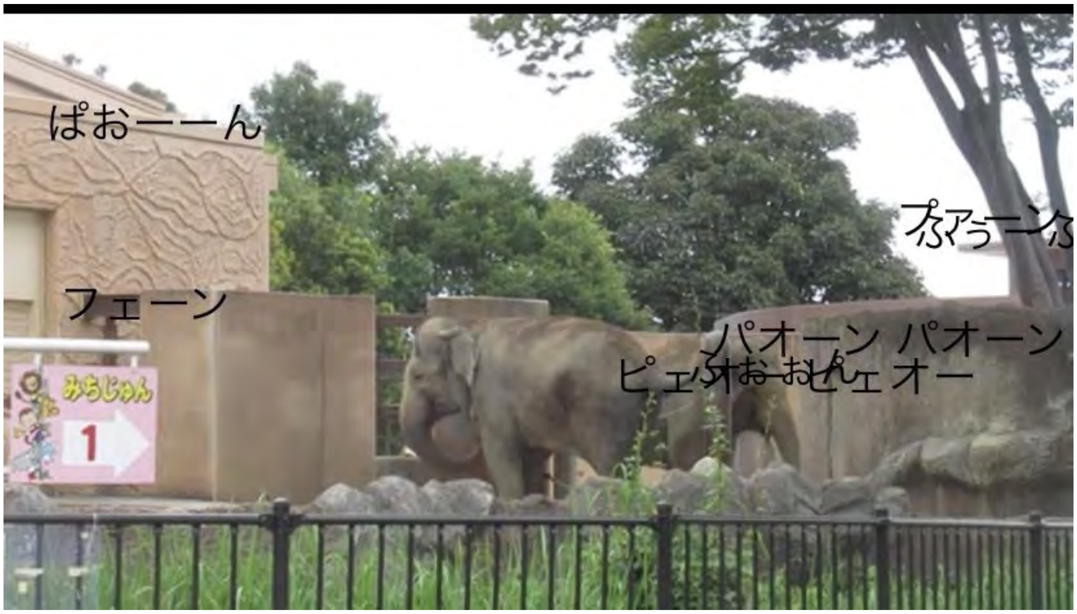


Fig. 1. Snapshot of Video with Superimposed Creative Onomatopoeia.

Fig. 2. System Conceptual Diagram.

Discussion

We conducted a survey to determine whether people were aware of the differences in the way people hear and perceive sounds, whether we could stimulate their interest in sounds and animals, and whether we could help them develop their listening skills. Fifteen DHH university students were asked to use the prototype of the system and complete an evaluation questionnaire. All the college students who participated in the evaluation questionnaire had hearing levels above 60 dB and we did not ask about personal information regarding their gender or detailed hearing. The questionnaire was administered to each student in a general classroom setting, in which they were given a tablet device and asked to freely manipulate it to play a video. Table 2 shows the results.

Table 2. The results of the survey.

Question	**Strong Agreement**	**Agreement**	**Neutral**	**Disagreement**	**Strong Disagreement**
Do you think that each person hears sound differently when using this app?	5	7	3	0	0
Do you enjoy learning about animal sounds?	7	5	2	1	0
Does this app make you want to visit the zoo?	3	2	6	1	3
Does this app help you develop your hearing?	1	4	5	2	3

The difference in hearing was rated positively by 67% (12/15) of the respondents. The enjoyment of learning about the animal sounds through the application was positively evaluated by 67% (12/15) of the respondents. Respondents were requested to give their comments, and the following were obtained:

- It is fun to visualize the calls as onomatopoeia
- I can see the text and understand what they are saying
- I thought it was a good teaching tool for children who need visual information because I have no auditory function to see and learn.
- There is an onomatopoeia that says it sounded like this, and I know that the way people hear it is certainly different.
- I thought the design was nice and child-friendly.

However, mixed positive, intermediate, and negative evaluations were observed of the system’s ability to encourage zoo visits as well as its listening comprehension training. In order to train listening skills, we believe it is necessary to provide a usage manual, combine the system with vibration speakers, or ask the students to listen in a quiet environment.

Conclusions

In this study we developed a system that visualizes animal sounds as creative onomatopoeia allowing people to imagine animal sounds as they hear them without being bound to conventional sound descriptions. The system was designed to allow everyone to accept and enjoy their own way of hearing. Although we believe that we were able to communicate the differences in hearing, we were unable to develop an interest in animals and initiate a visit to a zoo. In the future, we would like to investigate the relationship between onomatopoeia and DHH children, investigate global approaches, and further discuss onomatopoeia and hearing impairment to aid the acceptance of unique ways of hearing and interest in sounds.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number 21K18482.

Works Cited

DCASE Community. dcase.community/community_info

Ekimatope. https://japannews.yomiuri.co.jp/society/general-news/20220619-39082/

Hashizume, S., Sakamoto, S., Suzuki, K., & Ochiai, Y. (2018). “LIVEJACKET: Wearable Music Experience Device with Multiple Speakers. Distributed, Ambient and Pervasive Interactions: Understanding Humans.” DAPI. 2018. Lecture Notes in Computer Science, 10921. Springer. doi.org/10.1007/978-3-319-91125-0_30

MEXT. “A Guide to Education for the Deaf and Hard of Hearing.” Special Education System, Ministry of Education, Culture, Sports, Science and Technology, Mar. 2020, https://www.mext.go.jp/content/20200324-mxt_tokubetu02-100002897_003.pdf.

Mylar Ballon. https://insheepsclothinghifi.com/listening-bob-lichtenberg-stereophile/

Ontenna. ja-jp.facebook.com/ontenna/

Ramkalawan T. W. and Davis A. C. “The Effects of Hearing Loss and Age of Intervention on Some Language Metrics in Young Hearing-Impaired Children.” British Journal of Audiology, vol. 26, 2, 1992, pp. 97–107, doi:10.3109/03005369209077877.

Sound Display. https://www.tdi.co.jp/sound-display/

SoundHug. https://pixiedusttech.com/soundhug/en/index.html?language=en

Emergency Communication for People with Disabilities

Raeda Anderson[1,2], John Morris[1], Delaney Cowart[1], Ben Lippincott[1]
Shepherd Center[1] and Georgia State University, Department of Sociology[2]
raeda.anderson@shepherd.org; john.morris@shepherd.org;
delaney.cowart@shepherd.org; ben.lippincott@shepherd.org

Abstract

Communication during public emergencies is critical for comfort, security and survival. This is especially true for people with disabilities, who are considered a particularly vulnerable population. Understanding patterns of communication by people with disabilities - sources of emergency related information, communication partners, and modes of communication - can assist emergency response planners and also help keep people connected and secure. This paper presents survey research results from the "2021 National Test of the Emergency Alert System and Wireless Emergency Alerts", a self-administered online survey of 212 members of the Accessibility User Research Collective (AURC), in August of 2021. The most common source of information during an emergency for people with disabilities is local news (69.3%) followed by a weather or news mobile app (57.5%), word of mouth (42.9%), radio (40.6%), social media (38.7%), and national news (32.1%). The primary contact person during an emergency is a person's spouse or romantic partner (43.0%). Secondary contacts include family members (48.6%), friends (32.5%), and parents (25.5%) with significant differences between disability groups. Primary and secondary contacts are reached most commonly though text messages (63.7%), phone calls (54.7%), in person conversations (30.7%), and social media (20.3%) with some variation between disability groups.

Keywords

Emergency, communication, information, disability.

Introduction

Communication during public emergencies is of paramount importance to ensure safety, share resources, and disseminate information (Lien et. al np, Salamon np). Extensive analysis of communication has established that communicating via smart phones with social media (Beneito-Montagut, et al np, Panagiotopoulos et al. 87, Ma and Yated 110) and phone calls (Mythili and Shalini 392) are often primary forms of communication during an emergency. The importance of effective communication for people with disabilities has long been established (Davidson et al. 330, Hoffman et al. np, Khetarpal 99). More specifically, research has confirmed the importance of communication for people with disabilities to manage their day-to-day lives, in handling medical needs (Anderson et al. 266, Khetarpal 625, Pransky et al 626, Usmanov et al. 17), leisure (Cherney et al. 9), learning (Noens et al. 199), and mental health (Anderson et al. 140).

Fewer studies, however, have examined how people with disabilities communicate during emergencies, and nearly all of that scholarship focuses on personal medical emergencies, rather than widespread emergencies such as extreme weather (Hans and Mohanty 120, Queshem et al. np, Twigg et al. 249). Of the scholarship that examines communication during emergencies, communication is focused via a single platform of communication of social media (Morris et al. 567, Kent and Ellis 419) or text messages (Baker and Mitchell). While these works are important, they do not examine the multiple strands of communication across modes, that are likely common during emergencies for everyone, including people with disabilities. Additionally, given the substantive effort of Federal Emergency Management Agency (FEMA) to make WEA and EAS messaging accessible (Alerting People with Disabilities Across Functional Needs np), updated scholarship examining communication patterns and modes as well as data sources during an emergency are of paramount importance. Thus, this study sets out to examine communication patterns, different communication modes, and information sources for people with disabilities during an emergency.

Methods

Data and Measures

Data were collected via a self-administered online survey titled “2021 National Test of the Emergency Alert System and Wireless Emergency Alerts” in August of 2021. Study respondents are all members of the Accessibility User Research Collective (AURC), formerly known as the Consumer Advisory Network (CAN), which is a panel of US adults with disabilities. A total of 212 AURC members completed the survey.

Disability and Demographics

Measures were calculated either from a single indicator within the survey, or a combination of indicators. Disability measures in this study include walking, fatigue and limited stamina, blind, deaf, difficulty speaking, anxiety, upper extremity limitations, and learning disability. Most disability indicators were compiled from single responses to “Do you have any of the following conditions or difficulties? (select all that apply)”: *walking* via “difficulty walking, standing, or climbing stairs”, *fatigue and limited stamina* via “difficulty with fatigue/limited stamina”, *blind* via “blindness (without usable vision)”, *deaf* via “deafness”, *difficulty speaking* via “difficulty speaking so people can understand you”, *anxiety* via “frequent worrying, nervousness, or anxiety”, and *upper extremity limitations* via “difficulty using your hands or fingers”. *Learning disability*, however, was generated by combining multiple response options to “Do you have any of the following conditions or difficulties? (select all that apply)”, where respondents selected “difficulty concentrating, remembering, or making decisions” and/or "difficulty thinking (learning, remembering, or concentrating)". Demographic measures were generated from various measures including "What is your gender?" for gender and "Which race/ethnicity best describes you? (Please choose only one.)" for racial minority status.

Emergency Communication and Technology

Emergency communication and technology were accounted for via a series of indicators. Sources of information were measured via "What sources of information would you use during a severe weather event? (check all that apply)" with the choices of "local news", "national news", "social media", "radio", "direct contact with family/friends/colleagues (e.g. text, call, email, talk)", "weather app/news", and "other (please specify)" as a write in option. The primary contact person during an emergency was measured via "Who is the FIRST person you would contact?" and then compiled into the following categories for analysis: *no one* measured via "No one, I would not contact anyone", *parent* via "parent or guardian", *spouse* via "spouse/girlfriend/boyfriend/partner", *family* via "sibling", "child" and, "other family member", *friend* via "friend", and *other* via "coworker", "caregiver or personal care assistant", "roommate", "neighbor", "member of faith organization/rabbi/priest/minister", "member of local community group (e.g. neighborhood association, bowling league, American Legion) ", and "other (please specify)" as a write in option. Similarly, the secondary contact is measured via "Who else would you contact? (check all that apply)" with the same response options and compilation from primary contact person. Communication mode was asked immediately following the primary and secondary contact reported and measured via "How would you contact them? (check all that apply)" with "phone call", "text", "social media (e.g. Facebook, Twitter, NextDoor)", "email", "talk in person", and "other (please specify)".

Analysis

Analysis includes descriptive statistics of all study variables and chi-square measure of association for relational analysis. All relational analysis employs pairwise deletion and $p<0.05$ to determine statistical significance (Frankfort-Nachmias and Guererro 2016).

Results

Disability and Demographic Measures

Table 1: Disability and Demographics of Study Sample.

Disability	**%**
Disability: Anxiety	18.9
Disability: Fatigue and limited stamina	18.9
Disability: Difficulty speaking	5.2
Disability: Walking	27.8
Disability: Learning	15.1
Disability: Deaf	14.2
Disability: Blind	25.9
Disability: Upper extremity limitations	18.4
Demographics: Female	50.8
Demographics: Racial Minority	22.0

Table 1 summarizes the disability and demographic characteristics of the study sample. Disability was measured across eight categories: learning disability 15.1%, anxiety 18.9%, difficulty speaking 5.2%, upper extremity limitations 18.4%, walking 27.8%, fatigue and limited stamina 18.9%, blind 25.9%, and deaf 14.2%. 50.8% of respondents were female, and 22.0% were racial minorities.

Sources of Information During Emergency by Disability Group.

Table 2a. Sources of Information during Emergency by Disability Type and Chi-square Results.

Disability	Local news: %	Local news: sig	National news: %	National news: sig	Social Media: %	Social Media: sig
Disability: Anxiety	72.5		37.5		55.0	*
Disability: Fatigue and limited stamina	80.0		37.5		45.0	
Disability: Difficulty speaking	72.7		36.4		54.5	
Disability: Walking	78.0	ł	35.6		44.1	
Disability: Learning	71.9		34.4		50.0	
Disability: Deaf	56.7		30.0		46.7	
Disability: Blind	92.7	***	47.3	**	43.6	
Disability: Upper extremity limitations	79.5		43.6	ł	51.3	ł

Table 2b. Sources of Information during Emergency by Disability Type and Chi-square Results.

ł p<0.10, *p<0.05, **p<0.01, ***p<0.001

Overall Percent of Sources of Information: Local news 69.3%, National news 32.1%, Social Media: 38.7%, Radio: 40.6%, Word of Mouth: 42.9%, Weather App/News: 57.5%.

Disability	Radio: %	Radio: sig	Word of Mouth: %	Word of Mouth: sig	Weather App/News: %	Weather App/News: sig
Disability: Anxiety	30.0		42.5		62.5	
Disability: Fatigue and limited stamina	37.5		55.0	ł	57.5	
Disability: Difficulty speaking	54.5		36.4		36.4	

Disability	Radio: %	Radio: sig	Word of Mouth: %	Word of Mouth: sig	Weather App/News: %	Weather App/News: sig
Disability: Walking	45.8		45.8		59.3	
Disability: Learning	34.4		46.9		53.1	
Disability: Deaf	13.3	**	33.3		63.3	
Disability: Blind	65.5	***	65.5	***	69.1	*
Disability: Upper extremity limitations	46.2		48.7		64.1	

Table 2a and 2b show the sources of information used by people with disabilities during an emergency with local news (69.3%) being the most common followed by a weather mobile app (57.9%), word of mouth (42.9%), radio (40.6%), social media (38.7%), and national news (32.1%). People who are blind have systematically different sources of information than other disability groups as 92.7% get information from local news, 47.3% ($p<0.01$) get information from national news, 65.5% ($p<0.001$), 65.5% ($p< .001$) get information from word of mouth and 69.1% ($p<.05$) get information from weather apps. Thus, people who are blind are significantly more likely to get information during emergencies through all communication sources other than social media. people who are deaf are less likely to get information from the radio at 13.3% ($p<.01$) compared to people with other types of disabilities. People with anxiety are more likely to get emergency information from social media at 55.0% ($p<0.05$) compared to 38.7% of all people with disabilities. There are some marginally significant differences with other disability groups and forms of communication that are indicated on Table 2.

Contact for Communication During Emergency by Disability Group

Table 3. Primary Contact during Emergency by Disability Type and Chi-square Results.

ł p<0.10, *p<0.05, **p<0.01, ***p<0.001

Overall Percent of Primary Contact: No one 8.9%, Parent 14.5%, Spouse 43.0%, Family 17.3%, Friend 5.0%, Other 11.2%.

Disability	No one	Parent	Spouse	Family	Friend	Other	Analysis: χ^2	Analysis: sig
Disability: Anxiety	5.9%	11.8%	41.2%	20.6%	2.9%	17.6%	2.889	0.717
Disability: Fatigue and limited stamina	5.7%	11.4%	45.7%	14.3%	2.9%	20.0%	4.539	0.475
Disability: Difficulty speaking	10.0%	10.0%	30.0%	20.0%	10.0%	20.0%	1.883	0.865
Disability: Walking	9.1%	18.2%	36.4%	16.4%	1.8%	18.2%	6.709	0.243
Disability: Learning	7.1%	14.3%	28.6%	28.6%	10.7%	10.7%	6.229	0.278
Disability: Deaf	3.6%	17.9%	46.4%	10.7%	7.1%	14.3%	2.832	0.726
Disability: Blind	7.8%	13.7%	43.1%	15.7%	7.8%	11.8%	1.384	0.926
Disability: Upper extremity limitations	11.4%	22.9%	40.0%	14.3%	2.9%	8.6%	3.376	0.642

Table 3 presents information on the primary contact for people with disabilities across disability types. The most common primary contact is spouse (43.0%) followed by family at 17.3%, parent at 14.5%, other at 11.2%, no one at 8.9%, and friend at 5.0%. While there is variation between primary contact points in an emergency, there are not statistically significant differences between primary contacts in an emergency and disability type. It is worth noting, however, that people with learning disabilities reported the primary contact for them being evenly split between spouse and other family members, while all other disability groups spouse is the single largest contact point for primary contact.

Table 4a. Secondary Contacts during Emergency by Disability Type and Chi-square Results.

Disability	**No one: %**	**No one: sig**	**Parent: %**	**Parent: sig**	**Spouse: %**	**Spouse: sig**
Disability: Anxiety	15.0		32.5		12.5	
Disability: Fatigue and limited stamina	10.0		30.0		10.0	
Disability: Difficulty speaking	0.0		18.2		9.1	
Disability: Walking	13.6		25.4		10.2	
Disability: Learning	15.6		37.5	ł	15.6	
Disability: Deaf	3.3		23.3		13.3	
Disability: Blind	16.4		32.7		16.4	
Disability: Upper extremity limitations	15.4		23.1		5.1	ł

Table 4b. Secondary Contacts during Emergency by Disability Type and Chi-square Results.

ł p<0.10, *p<0.05, **p<0.01, ***p<0.001

Overall Percent of Secondary Contact (select all that apply): No one 11.8%, Parent 25.5%, Spouse 13.2%, Family 48.6%, Friend 32.5%, Other 35.8%.

Disability	**Family: %**	**Family: sig**	**Friend: %**	**Friend: sig**	**Other: %**	**Other: sig**
Disability: Anxiety	50.0		37.5		40.0	
Disability: Fatigue and limited stamina	67.5	**	30.0		50.0	*
Disability: Difficulty speaking	45.5		45.5		45.5	
Disability: Walking	57.6		27.1		50.8	**

Disability	Family: %	Family: sig	Friend: %	Friend: sig	Other: %	Other: sig
Disability: Learning	43.8		21.9		40.6	
Disability: Deaf	63.3	ł	46.7	ł	40.0	
Disability: Blind	50.9		45.5	*	38.2	
Disability: Upper extremity limitations	53.8		25.6		56.4	**

Table 4a and 4b present information on all secondary contacts identified by people with disabilities across disability type. The most common secondary contact identified is a family member (48.6%) followed by others (35.8%), friends (32.5%), parents (25.5%), spouse (13.2%), and lastly, no secondary contacts (11.8%). Unlike primary contacts, there were significant differences in patterns of contacts across disability groups with the most notable patterns for respondents with fatigue and limited stamina reporting that they were more likely to contact their family ($p<0.01$) and others ($p<0.05$), compared to other disability groups. Additionally, people with walking limitations ($p<0.01$) and upper extremity limitations ($p<0.01$) also reported significantly higher likelihood of secondary contacts that are not one of the main groups and thus those in the 'Other' category.

Communication Mode During an Emergency by Disability Type

Table 5a. Communication Mode During an Emergency and Disability Type and Chi-square Results.

Disability	Phone Call: %	Phone Call: Sig	Text Message: %	Text Message: Sig	Social Media: %	Social Media: Sig
Disability: Anxiety	55.0		67.5		30.0	ł
Disability: Fatigue and limited stamina	60.0		70.0		32.5	*
Disability: Difficulty speaking	18.2	*	90.9	ł	18.2	
Disability: Walking	62.7		72.9	ł	25.4	

Disability	Phone Call: %	Phone Call: Sig	Text Message: %	Text Message: Sig	Social Media: %	Social Media: Sig
Disability: Learning	56.3		62.5		28.1	
Disability: Deaf	33.3	*	80.0	*	16.7	
Disability: Blind	72.7	**	74.5	ł	23.6	
Disability: Upper extremity limitations	61.5		69.2		28.2	

Table 5b. Communication Mode During an Emergency and Disability Type and Chi-square Results.

ł p<0.10, *p<0.05, **p<0.01, ***p<0.001

Overall Percent of Communication Mode: Phone Call 54.7%, Text Message 63.7%, Social Media: 20.3%, Email: 15.1%, Talk in Person: 30.7%, Other: 2.8%.

Disability	Email: %	Email: Sig	Talk in Person: %	Talk in Person: Sig	Other: %	Other: Sig
Disability: Anxiety	15.0		32.5		2.5	
Disability: Fatigue and limited stamina	15.0		32.5		7.5	*
Disability: Difficulty speaking	9.1		18.2		0.0	
Disability: Walking	11.9		39.0		3.4	
Disability: Learning	15.6		31.3		6.3	
Disability: Deaf	16.7		23.3		3.3	
Disability: Blind	23.6	*	41.8	*	1.8	
Disability: Upper extremity limitations	10.3		41.0		5.1	

Communication modes during an emergency by disability type are displayed in Table 5a and 5b across six communication modes with the most common being text messages (63.7%), followed by phone calls (54.7%), talking in person (30.7%), social media (20.3%), email

(15.1%), and other communication modes (2.8%). Similar to sources of information, people who are blind have the most notable differences from people with other types of disabilities, as they are more likely to call ($p<0.01$), email ($p<0.05$), and talk in person ($p<0.05$) to communicate during an emergency. Additionally, 33.0% of deaf respondents reported phone calls as a communication mode compared to 54.7% of people with other types of disabilities ($p<0.05$) and they reported 80.0% communicating via text messages compared to 63.7% of people with other types of disabilities ($p<0.05$). People with fatigue and limited stamina reported communicating via social media 11.8% more ($p<0.05$) during an emergency than people from all disability groups.

Conclusion

Accessibility, reliability, timeliness and trustworthiness of the communication channels and the information transmitted through them are critical concerns at all times, especially during public emergencies, which are disruptive, with rapidly changing conditions. News and information can be outright incorrect or grow stale quickly as the situation changes. Furthermore, having communications access to emergency management personnel, family, friends and neighbors can ease worry and help distribution of potentially scarce resources. Public emergencies, especially when they are prolonged as happened in the aftermath of Hurricane Katrina and Sandy, or during large wildfires, can disrupt access to medications, assistive technology, direct care workers (personal care assistants), healthcare providers and more, threatening the health and safety of individuals with disabilities and raising uncertainty and concern by friends and family.

Emergency communications varies across disability type. People with disabilities vary in their sources of information, secondary contacts, and communication modes during an emergency. Thus, when emergency response is planned, specific attention should be given to the sources of information and communication options available to people with disabilities to ensure even dissemination of information and effective communication.

Acknowledgments

This research is funded by The Rehabilitation Engineering Research Center for Information and Communications Technology Access (LiveWell RERC), a 5-year grant from the National Institute on Disability, Independent Living and Rehabilitation Research (NIDILRR) in the U.S. Department of Health and Human Services (grant number 90RE5028). The opinions contained in this publication are those of the LiveWell RERC and do not necessarily reflect those of the U.S. Department of Health and Human Services or NIDILRR.

Thank you to all study participants. Without your willingness to participate in research and your insights, our research would not be possible. Additionally, thank you to Sarah Mosely, Matthew Cruea, and George Usmanov for your assistance in supporting this study's data collection, development, and finalization.

Works Cited

Alerting People with Disabilities and Access and Functional Needs. March 1, 2022. https://www.fema.gov/emergency-managers/practitioners/integrated-public-alert-warning-system/public/alerting-people-disabilities Accessed November 15, 2022

Alper, Meryl. *Giving voice: Mobile communication, disability, and inequality*. MIT Press, 2017.

Anderson, Raeda K., et al. “Disability, Technology Based Communication, and Loneliness.” *The Journal on Technology and Persons with Disabilities* 140 (2021).

Anderson, Raeda, George Usmanov, and Nicole Thompson. “People with Disabilities Online Engagement During COVID-19.” (2022).

Beneito-Montagut, Roser, et al. “Governmental social media use for emergency communication.” *ISCRAM*. 2013.

Bennett, DeeDee, Paul MA Baker, and Helena Mitchell. “New media and accessible emergency communications: A United States-based meta-analysis.” *Disability and Social Media*. Routledge, 2016. 141-152.

Boon, Helen J., et al. “School Disaster Planning for Children with Disabilities: A Critical Review of the Literature.” *International Journal of Special Education* 26.3 (2011): 223-237.

Cherney, James L., Kurt Lindemann, and Marie Hardin. “Research in communication, disability, and sport.” *Communication & Sport* 3.1 (2015): 8-26.

Davidson, Bronwyn, et al. “Social participation for older people with aphasia: The impact of communication disability on friendships.” *Topics in stroke rehabilitation* 15.4 (2008): 325-340.

Frankfort-Nachmias, Chava, Anna Leon-Guerrero, and Georgiann Davis. *Social statistics for a diverse society*. Sage Publications, 2019.

Goggin, Gerard, and Katie Ellis. “Disability, communication, and life itself in the COVID-19 pandemic.” *Health sociology review* 29.2 (2020): 168-176.

Gordon, Clare, Caroline Ellis-Hill, and Ann Ashburn. “The use of conversational analysis: nurse-patient interaction in communication disability after stroke.” *Journal of advanced nursing* 65.3 (2009): 544-553.

Hans, Ashok, and Reena Mohanty. “Disasters, disability and technology.” *Development* 49.4 (2006): 119-122.

Hoffman, Jeanne M., et al. “Effect of communication disability on satisfaction with health care.” (2005).

Kent, M., & Ellis, K. (2015). People with disability and new disaster communications: access and the social media mash-up. *Disability & Society*, *30*(3), 419-431.

Khetarpal, Abha. “Information and communication technology (ICT) and disability.” *Review of market integration* 6.1 (2014): 96-113.Jamwal, Rebecca, et al. “Smart home and communication technology for people with disability: a scoping review.” *Disability and Rehabilitation: Assistive Technology* 17.6 (2022): 624-644.

Khetarpal, Abha. “Information and communication technology (ICT) and disability.” *Review of market integration* 6.1 (2014): 96-113.

Lien, Yao-Nan, Li-Cheng Chi, and Yuh-Sheng Shaw. “A walkie-talkie-like emergency communication system for catastrophic natural disasters.” *2009 10th International Symposium on Pervasive Systems, Algorithms, and Networks*. IEEE, 2009.

Ma, Xin, and Justin Yates. “Optimizing social media message dissemination problem for emergency communication.” *Computers & Industrial Engineering* 78 (2014): 107-126.

Morris, John T., James L. Mueller, and Michael L. Jones. “Use of social media during public emergencies by people with disabilities.” *Western journal of emergency medicine* 15.5 (2014): 567.

Mythili, S., and E. Shalini. “A comparative study of smart phone emergency applications for disaster management.” *International Research Journal of Engineering and Technology (IRJET)* 3.11 (2016): 392-395.

Noens, Ilse, and Ina van Berckelaer-Onnes. “Making sense in a fragmentary world: Communication in people with autism and learning disability.” *Autism* 8.2 (2004): 197-218.

Panagiotopoulos, Panos, et al. “Social media in emergency management: Twitter as a tool for communicating risks to the public.” *Technological Forecasting and Social Change* 111 (2016): 86-96.

Pransky, Glenn S., et al. “Disability prevention and communication among workers, physicians, employers, and insurers — current models and opportunities for improvement.” *Disability and rehabilitation* 26.11 (2004): 625-634.

Qushem, Umar Bin, Abdul Rahman Bin Ahmad Dahlan, and Ahmad Shahril Bin Mohd Ghani. “My emergency assistant device: A conceptual solution in enhancing the quality of life for the disabled and elderly.” *2016 6th International Conference on Information and Communication Technology for The Muslim World (ICT4M)*. IEEE, 2016.

Salamon, Margaret Klein. “Leading the public into emergency mode.” *The Climate Mobilization* (2019).

Spence, Patric R et al. “Media use and information needs of the disabled during a natural disaster.” *Journal of health care for the poor and underserved* vol. 18,2 (2007): 394-404. doi:10.1353/hpu.2007.0047

Twigg, John, et al. “Disability and public shelter in emergencies.” *Environmental Hazards* 10.3-4 (2011): 248-261.

Usmanov, George S., Eric R. Wright, and Raeda K. Anderson. “Treatment Network Typologies and the Working Alliance of Clients with Serious Mental Illness.” *Society and Mental Health* 12.1 (2022): 17-31.

THE JOURNAL ON TECHNOLOGY AND PERSONS WITH DISABILITIES

Disability and Telehealth: Healthcare Access and Motivations

Raeda Anderson[1,2], Suveyda Karakaya[1], Elorm Adzadi[1], J'Lyn Martin[1], John Morris[1]
Shepherd Center
Georgia State University, Department of Sociology
raeda.anderson@shepherd.org, suveyda.karakaya@shepherd.org, elorm.adzadi@shepherd.org, j'lyn.martin@shepherd.org, john.morris@shepherd.org

Abstract

There has been a rapid increase in telehealth adoption in recent years driven by the 2020 COVID-19 pandemic yet, there is much we do not know about the experiences with telehealth by people with disabilities. It is also unclear how the increasing shift to telehealth alters healthcare access for socially and economically disadvantaged groups. This paper analyzes survey research data on telehealth use and compares telehealth use patterns across groups with different types of disabilities. Convenience sampling was used to collect data via an online survey from 326 adults with various disabilities. Participants were recruited through the Accessibility User Research Collective (AURC), a national network of people with disabilities in the United States developed and maintained by researchers at Shepherd Center, a rehabilitation hospital in Atlanta, Georgia. Findings from this study can help inform policy related to telehealth and telehealth access.

Keywords

Telehealth, disability, COVID-19, technology.

Introduction

Since the onset of the COVID-19 pandemic, there has been a rapid increase in telehealth adoption. While telehealth use has become more mainstream, facilitators and barriers to telehealth access for underserved communities, especially for people with disabilities, remain underexplored. There are 61 million adults with disabilities living in the United States which is about 26% of the adult population (Centers for Disease Control and Prevention, 2020). People with disabilities experience significant disparities in access to health care (Friedman and VanPuymbrouck 2021; Valdez et al. 2021). It is important to understand how the shift to telehealth impacts healthcare access for people with disabilities. This study explores the relationship between disability type and other demographic variables and patterns of telehealth use and motivations for using or not using telehealth. Findings from this study will help to inform policies to achieve equitable telehealth access.

For this study, we define telehealth as healthcare or a consultation received remotely from a clinical professional by phone or video call. There is no consensus among scholars regarding the overall impact of telehealth adoption on healthcare access for people with disabilities. However, telehealth has the potential to increase the accessibility of healthcare. Some case studies and anecdotal evidence support the argument that telehealth could benefit people with disabilities as well as the general population due to efficiency, convenience, and reduced barriers to healthcare (Lee et al. 2020; Sechrist 2018). Compared to in-person visits, one unique benefit of telehealth is the ability to connect patients and doctors at a distance and eliminate transportation barriers. This could be appealing especially for people with mobility/dexterity disabilities. Indeed, Friedman and VanPuymbrouck (2021) found that people with mobility disabilities were significantly more likely to use telehealth during the second year of the pandemic compared to people with other disabilities. They also found that people with cognitive disabilities were more likely to use telehealth but in contrast, people with hearing disabilities were less likely to use telehealth. Conversely, skeptics argue that telehealth is only accessible to individuals who can afford the internet and modern electronic devices which could increase health inequalities; the digital divide remains an obstacle to equitable telehealth adoption (Chang, et al. 2021; Chunara et al. 2021; Haynes et al. 2021; Valdez et al. 2021).

Additionally, communicating via video or phone call could be challenging for people with certain types of disabilities such as those with speech disabilities, people who are deaf or hard of hearing, blind or low vision, and who have intellectual disabilities (Lee et al. 2020; Mitchell 2018; Valdez et al. 2021). Beyond disability-related barriers, some people may also experience technological constraints or concerns. These may include limited access to fast and/or reliable internet in rural areas, compatibility with external assistive technology devices, and unsecure networks containing sensitive health information (Lee et al. 2020; Valdez et al. 2021). Many people with disabilities live in rural and low-income areas where broadband internet is not available or reliable. Uneven adoption of web accessibility standards by telemedicine platforms and a dearth of regulations under the ADA that specifically address web accessibility also present barriers (Annaswamy et al., 2020). These unique challenges need to be addressed to better provide telehealth access for people with disabilities.

With mixed findings in the current literature, continued research is needed to better understand how telehealth use differs across people with disabilities. This study analyzes survey research data collected by the Accessibility User Research Collective at Shepherd Center to explore telehealth use patterns among people with disabilities. Specifically, the survey asked respondents about the frequency of telehealth use, their preference for telehealth vs. in-person healthcare, their motivations to prefer telehealth, devices and assistive technology used to access telehealth services, and the type of healthcare received via telehealth.

Methods

Data were collected from June 16 to July 5, 2022, via a self-administered online survey using convenience sampling. Our sample of participants was recruited via the Accessibility User Research Collective (AURC), a network of 1,305 adults with disabilities across the United States of America. This convenience sample was used as AURC members regularly participate as a panel in research surrounding disability and technology. Additionally, nationally representative samples of people with disabilities are rare and cost prohibitive for many research studies, including this scholarship. A total of 326 AURC members completed the survey, representing a response rate of 24.98%. Analysis includes descriptive statistics, measures of association, and logistic regression analysis to explore telehealth use patterns across different types of disabilities using pairwise deletion with a standard $p<0.05$ for statistical significance (Frankfort-Nachmias and Guererro 2016).

Analysis

Table 1 contains demographic and disability measures for study participants overall and by their telehealth use status. 79.5% of study participants had used telehealth. Disability types varied widely with 25.2% with a learning disability, 27.2% with anxiety, 11.9% with difficulty speaking, 24.8% with upper extremity limitations, 34.7% with difficulty walking, 24.3% with fatigue and limited stamina, 20.5% blind, and 17.8% deaf. Telehealth use was common across all disability groups with the highest telehealth use among study participants with upper extremity limitations (94.3%) followed by walking (90.9%), learning disability (87.9%), fatigue and limited stamina (87.5%), anxiety (85.9%), difficulty speaking (85.7%), deaf (77.8%), and blind

(69.0%). For demographic measures, 52.8% of respondents were female and 76.1% were white. Average age was 51.09 years old (SD=14.21).

Table 1. Disability and Demographic Measures by Telehealth Use.

Characteristic	**Overall %**	**Does not use Telehealth %**	**Uses Telehealth %**
Use telehealth	79.5	0.0	100.0
Disability: Learning disability	25.2	12.1	87.9
Disability: Anxiety	27.2	14.1	85.9
Disability: Difficulty speaking	11.9	14.3	85.7
Disability: Upper extremity limitations	24.8	5.7	94.3
Disability: Walking	34.7	9.1	90.9
Disability: Fatigue and limited stamina	24.3	12.5	87.5
Disability: Blind	20.5	31.0	69.0
Disability: Deaf	17.8	22.2	77.8
Demographic: Female	52.8	17.4	82.6
Demographic: White	76.1	21.8	78.2
Demographic: Age	mean (SD): 51.09 (14.21)	mean (SD): 52.80 (15.29)	mean (SD): 50.54 (13.88)

Telehealth Use

Table 2a, 2b, and 2c present analysis of telehealth use by disability type while controlling for factors that have been established to affect telehealth use using odds ratio logistic regression models. Respondents with walking limitations are 3.85 times more likely to use telehealth, compared to other types of disabilities (OR=3.85, $p<0.01$). Similarly, people with upper extremity limitations were 5.50 times more likely to use telehealth (OR= 5.50, $p<0.01$) and people with fatigue and limited stamina were marginally more likely to use telehealth (OR=2.20, p<0.10). Conversely, blind respondents were 0.42 times less likely to use telehealth than people with other types of disabilities (OR=0.42, p<0.05). Respondents who have anxiety (OR=1.77, p>0.10), difficulty speaking (OR=1.82, p>0.10), learning disabilities (OR=2.12, p>0.10), and are deaf (OR=0.87, p>0.10) were not significantly more or less likely to have had a telehealth appointment.

Table 2a. Telehealth use by disability type and control variables, logistic regression with odds ratio (OR).

Characteristic	**1 Control: OR**	**1 Control: sig**	**2 Anxiety: OR**	**2 Anxiety: sig**	**3 Speaking: OR**	**3 Speaking: sig**
Disability: Anxiety			1.77			
Disability: Difficulty speaking					1.82	
Disability: Walking						
Disability: Fatigue & Stamina						
Disability: Learning disability						
Disability: Deaf						
Disability: Blind						
Disability: Upper extremity						
Control: Rural	0.48		0.46		0.52	
Control: Speak English in Home	3.33	*	3.17	*	3.56	*
Control: Education	1.30	ł	1.32	*	1.31	ł
Control: Income	1.10		1.09		1.09	
Control: Disability Lifelong	0.56	ł	0.54	ł	0.55	ł

Characteristic	1 Control: OR	1 Control: sig	2 Anxiety: OR	2 Anxiety: sig	3 Speaking: OR	3 Speaking: sig
Control: Senior	2.47	*	2.29	*	2.59	*
Model fit: constant	0.21		0.19		0.169	ł
Model fit: r^2	0.0844		0.0917		0.0889	

Table 2b. Telehealth use by disability type and control variables, logistic regression with odds ratio (OR).

Characteristic	4 Walking: OR	4 Walking: sig	5 Fatigue: OR	5 Fatigue: sig	6 Learning: OR	6 Learning: sig
Disability: Anxiety						
Disability: Difficulty speaking						
Disability: Walking	3.85	**				
Disability: Fatigue & Stamina			2.22	ł		
Disability: Learning disability					2.12	
Disability: Deaf						
Disability: Blind						
Disability: Upper extremity						
Control: Rural	0.43		0.46		0.46	
Control: Speak English in Home	3.28	ł	3.33	*	3.66	*
Control: Education	1.29	ł	1.31	ł	1.31	*
Control: Income	1.13		1.09		1.10	
Control: Disability Lifelong	0.69		0.58		0.58	
Control: Senior	2.84	*	2.70	*	2.25	*

Characteristic	4 Walking: OR	4 Walking: sig	5 Fatigue: OR	5 Fatigue: sig	6 Learning: OR	6 Learning: sig
Model fit: constant	0.11	*	0.16	ł	0.16	ł
Model fit: r^2	0.1270		0.0980		0.0960	

Table 2c. Telehealth use by disability type and control variables, logistic regression with odds ratio (OR).

Ł p<0.10, *p<0.05, **p<0.01, ***p<0.001.

Characteristic	7 Deaf: OR	7 Deaf: sig	8 Blind: OR	8 Blind: sig	9 U Extremity: OR	9 U Extremity: sig
Disability: Anxiety						
Disability: Difficulty speaking						
Disability: Walking						
Disability: Fatigue & Stamina						
Disability: Learning disability						
Disability: Deaf	0.87					
Disability: Blind			0.42	*		
Disability: Upper extremity					5.50	**
Control: Rural	0.57		0.42		0.58	
Control: Speak English in Home	3.74	*	3.35	*	3.90	*
Control: Education	1.32	*	1.34	*	1.31	ł
Control: Income	1.10		1.11		1.13	
Control: Disability Lifelong	0.58		0.64		0.64	
Control: Senior	2.61	*	2.75	*	2.47	*

Characteristic	7 Deaf: OR	7 Deaf: sig	8 Blind: OR	8 Blind: sig	9 U Extremity: OR	9 U Extremity: sig
Model fit: constant	0.16	ł	0.18		0.11	*
Model fit: r^2	0.0861		0.1104		0.1262	

Healthcare Received Through Telehealth

Table 3a and 3b summarize types of healthcare services received through telehealth by disability type. Blind participants were more likely to have had annual checkups via telehealth (77.5%, $p<0.05$). People with fatigue and limited stamina (67.3%, $p<0.01$) and difficulty walking (62.9%, $p<0.01$) were more likely to have had appointments with their specialists via telehealth. People with learning disabilities reported having a 24/7 telehealth consultation more frequently (21.6%, $p<0.05$). People with anxiety (60.0%, $p<0.001$) and people with learning disabilities (56/9%, $p<0.001$) were more likely to have had counseling and therapy via telehealth.

Table 3a. Healthcare Received through telehealth appointments and disability type, chi-square.

Disability	Annual Check-Ups: %	Annual Check-Ups: Sig	Specialist: %	Specialist: Sig	24/7 health consultation: %	24/7 health consultation: Sig	Illness & Injury: %	Illness & Injury: Sig
Disability: Anxiety	60.0		47.3		18.2		25.5	
Disability: Difficulty speaking	63.3		67.3	**	20.4		20.4	
Disability: Walking	66.7		62.5		16.7		20.8	
Disability: Fatigue & Stamina	65.7		62.9	**	11.4		18.6	
Disability: Learning disability	72.5		54.9		21.6	*	21.6	

Disability	Annual Check-Ups: %	Annual Check-Ups: Sig	Specialist: %	Specialist: Sig	24/7 health consultation: %	24/7 health consultation: Sig	Illness & Injury: %	Illness & Injury: Sig
Disability: Deaf	51.4		45.7		13.2		20.0	
Disability: Blind	77.5	*	37.5	ł	10.0		17.5	
Disability: Upper extremity	66.0		56.0		10.0		20.0	

Table 3b. Healthcare Received through telehealth appointments and disability type, chi-square.

Disability	Physical Therapy: %	Physical Therapy: Sig	Occupational Therapy: %	Occupational Therapy: Sig	Speech Therapy: %	Speech Therapy: Sig	Counseling & Therapy: %	Counseling & Therapy: Sig
Disability: Anxiety	3.6		3.6		3.6		60.0	***
Disability: Difficulty speaking	8.2		6.1		4.1		38.8	
Disability: Walking	8.3		4.2		0.0		37.5	
Disability: Fatigue & Stamina	5.7		5.7	*	5.7	ł	28.6	
Disability: Learning disability	3.9		3.9		5.9		56.9	***
Disability: Deaf	5.7		0.0		2.9		34.3	
Disability: Blind	2.5		0.0		5.0		32.5	
Disability: Upper extremity	2.0		4.0		4.0		26.0	

Ł p<0.10, *p<0.05, **p<0.01, ***p<0.001

Overall Percent of Use via Telehealth: Annual Check-Ups 63.4%, Specialist Appointment 50.0%, 24/7 health consultation 13.4%, Illness and Injury

18.8%, Physical Therapy 5.0%, Occupational Therapy 2.5%, Speech Therapy 3.0%, Counseling and therapy 32.2%

Motivations for Using Telehealth

Table 4a and 4b summarize motivations for using telehealth by disability type. Concerns about contracting COVID-19 were the most common motivation (52.0%) with a marginally higher percentage for people with upper extremity limitations at 64.0% ($p<0.10$) and significantly lower for deaf respondents at 34.3% ($p<0.05$). 38.1% identified saving time as a motivation to use telehealth with a higher prevalence for people with learning disabilities (51.0%, $p<0.05$).

Table 4a. Telehealth motivations by disability type, chi-square.

Disability	**COVID-19: %**	**COVID-19: Sig**	**Distance: %**	**Distance: Sig**	**Time: %**	**Time: Sig**	**Cost: %**	**Cost: Sig**
Disability: Anxiety	47.3		38.2	ɫ	40.0		21.8	
Disability: Difficulty speaking	59.2		42.9	*	40.8		28.6	*
Disability: Walking	37.5		41.7		50.0		33.3	ɫ
Disability: Fatigue & Stamina	58.6		37.1	ɫ	32.9		21.4	
Disability: Learning disability	58.8		45.1	**	51.0	*	21.6	
Disability: Deaf	34.3	*	14.3	*	31.4		22.9	
Disability: Blind	57.5		15.0	*	32.5		15.0	
Disability: Upper extremity	64.0	ɫ	36.0		34.0		22.0	

Table 4b. Telehealth motivations by disability type, chi-square.

Disability	**Transportation: %**	**Transportation: Sig**	**Preference: %**	**Preference: Sig**	**Access: %**	**Access: Sig**	**Ins.: %**	**Ins.: Sig**
Disability: Anxiety	38.2		34.5	*	30.9	*	7.3	

Disability	Transportation: %	Transportation: Sig	Preference: %	Preference: Sig	Access: %	Access: Sig	Ins.: %	Ins.: Sig
Disability: Difficulty speaking	44.9	*	42.9	***	32.7	*	14.3	**
Disability: Walking	37.5		33.3		33.3	ł	8.3	
Disability: Fatigue & Stamina	44.3	*	32.9	*	30.0	*	11.4	*
Disability: Learning disability	41.2		37.3	**	29.4	ł	7.8	
Disability: Deaf	11.4	**	17.1		25.7		5.7	
Disability: Blind	45.0	ł	17.5		5.0	**	5.0	
Disability: Upper extremity	48.0	**	34.0	*	28.0		6.0	

Ł $p<0.10$, *$p<0.05$, **$p<0.01$, ***$p<0.001$

Overall percent of telehealth motivations: COVID-19 52.0%, Distance 28.7%, Time 38.1%, Cost 18.8%, Transportation 32.7%, Preference 23.3%, Access 20.3%, Insurance 5.9%

Conclusion

The survey research data presented here shows some significantly different patterns of overall telehealth use, types of healthcare received through telehealth, and motivations for using telehealth across disability types. Annual checkups and specialist appointments were the most common types of telehealth services received. The most prevalent motivations for using telehealth were the reduction of exposure to COVID-19 and saving time.

These results address high level questions of who uses telehealth services among people with disabilities, what services they use and why. Our survey on telehealth and disability provides additional opportunity for more detailed analysis of attitudes and activity related to telehealth use by people with disabilities and the intersectionality of age, economic status, technology adoption and racial/ethnic identity. Such intersectional analysis is critical to understanding the experiences of the "whole person" and ensuring that they receive timely, effective and efficient healthcare services delivered the way the patients with disabilities prefer and need.

Acknowledgments

This research is funded by The Rehabilitation Engineering Research Center for Information and Communications Technology Access (LiveWell RERC), a 5-year grant from the National Institute on Disability, Independent Living and Rehabilitation Research (NIDILRR) in the U.S. Department of Health and Human Services (grant number 90RE5028). The opinions contained in this publication are those of the LiveWell RERC and do not necessarily reflect those of the U.S. Department of Health and Human Services or NIDILRR. Additionally, two authors were funded to participate in this study through the Georgia Research Alliance Summer Scholar Program.

Thank you to all study participants. Without your willingness to participate in research and your insights, our research would not be possible. Additionally, thank you to Ben Lippincott, Delaney Cowart, Sarah Mosely, Cathy Furbish, and Matthew Cruea for your assistance in supporting this study's data collection, development, and finalization.

Works Cited

Annaswamy, Thiru M., Monica Verduzco-Gutierrez, and Lex Frieden. "Telemedicine barriers and challenges for persons with disabilities: COVID-19 and beyond." *Disability and health journal* 13.4 (2020): 100973.

Centers for Disease Control and Prevention, 2020. Disability Impacts All of Us Infographic | CDC. Accessed on July 5th, 2022.

Chang, Ji E., et al. "Rapid transition to telehealth and the digital divide: implications for primary care access and equity in a post-COVID era." *The Milbank Quarterly* 99.2 (2021): 340-368.

Chunara, Rumi, Yuan Zhao, Ji Chen, Katharine Lawrence, Paul A Testa, Oded Nov & Devin M Mann. “Telemedicine and Healthcare Disparities: A Cohort Study in a Large Healthcare System in New York City during COVID-19”. *Journal of the American Medical Informatics Association* vol. 28, no. 1, 2021, pp. 33-41.

Frankfort-Nachmias, Chava, and Anna Leon-Guerrero. *Social statistics for a diverse society*. Sage Publications, 2016.

Friedman, Carli, and Laura VanPuymbrouck. “Telehealth Use by Persons with Disabilities During the COVID-19 Pandemic.” *International Journal of Telerehabilitation* vol. 13, no. 2, 2021, pp. 1-14.

Haynes, Norrisa, Agnes Ezekwesili, Kathryn Nunes, Edvard Gumbs, Monique Haynes & JaBaris Swain. “Can You See My Screen? Addressing Racial and Ethnic Disparities in Telehealth.” *Current Cardiovascular Risk Reports*, vol. 15, no. 12, 2021, pp. 15-23.

Lee, Nicol Turner, Jack Karsten, and Jordan Roberts. “Removing Regulatory Barriers to Telehealth Before and After COVID-19.” *Brookings Institution*, 2020, pp. 1-23.

Mitchell, Uchechi A., Perla G. Chebli, Laurie Ruggiero, and Naoko Muramatsu. “The Digital Divide in Health-Related Technology Use: The Significance of Race/Ethnicity.” *The Gerontologist* vol. 59, no. 1, 2019, pp. 6-14.

Sechrist, Samantha, Sarah Lavoie, Cria-May Khong, Benjamin Dirlikov, and Kazuko Shem the COVID-19 Pandemic.” *International Journal of Telerehabilitation* vol. 13, no. 2, 2021, pp. 1-14.

Valdez, Rupa S, Courtney C Rogers, Henry Claypool, Lucy Trieshmann, Olivia Frye, Claire Wellbeloved-Stone, Poorna Kushalnagar. “Ensuring Full Participation of People with Disabilities in an Era of Telehealth.” *Journal of the American Medical Informatics Association*, vol. 28., no. 2, 2021, pp. 389-392.

Comparative Survey of Blind, Low-Vision and Sighted Workers on Crowdsourcing Platforms

THE JOURNAL ON TECHNOLOGY AND PERSONS WITH DISABILITIES

Ying Zhong, Makoto Kobayashi, Masaki Matsubara and Atsuyuki Morishima
University of Tsukuba, Tsukuba University of Technology
zhong.ying.wp@alumni.tsukuba.ac.jp, koba@cs.k.tsukuba-tech.ac.jp, masaki@slis.tsukuba.ac.jp, morishima-office@ml.cc.tsukuba.ac.jp

Abstract

Crowdsourcing offers a new form of work that makes it easy for people to work from home on their computers. The completion of tasks, however, can be difficult for those who have certain disabilities, such as visual impairment. There is a lack of research on comparing visually impaired and sighted crowd workers on non-English crowdsourcing platforms. This research aims to determine the best ways to support visually impaired workers with AI and change the modality to improve the performance of visually impaired crowd workers, especially screen reader users. This study used a questionnaire with 10 visually impaired and 20 sighted crowd workers to investigate the problems encountered by visually impaired people on Japanese crowdsourcing platforms. The key findings are as follows. (1) The number of task types for visually impaired workers is limited compared to the types of tasks finished by sighted workers on crowdsourcing platforms. (2) More accessible platforms and tasks for screen readers are crucial for visually impaired crowd workers. (3) The time taken by visually impaired crowd workers to perform tasks is a crucial problem. We also include suggestions for future work as part of this study and provide future directions.

Keywords

Crowdsourcing, Visually-Impaired Workers, Workers with Disabilities, Labor Platforms, Online Work, Remote Work, Gig Work.

Introduction

Crowdsourcing offers a new form of work that makes it easy for people to work from home on their computers. Crowdsourcing platforms are emerging in increasing numbers. These crowdsourcing platforms provide functionalities to match tasks with workers.

Crowd workers can choose the tasks they can perform on the crowdsourcing platform. Performing online tasks can however be problematic, for some people with disabilities, such as visually impaired people (Bourne et al., Zyskowski et al.).

In this study, an online survey was conducted to investigate the current status of crowdsourcing platforms and the problems encountered by visually impaired workers when performing these tasks. We also compared their working conditions with those of sighted crowd workers. A possible solution to increase the number of potential workers, including visually impaired workers, who can obtain new jobs in a crowdsourcing platform, has been provided herein.

The key findings of this paper are as follows:

- The number of task types for visually impaired workers is limited compared to the types of tasks finished by sighted workers on crowdsourcing platforms. This suggests that it will be beneficial to develop methods that can transform the tasks into those more accessible to visually impaired workers.
- Most visually impaired crowd workers are screen reader users. This suggests that visually impaired crowd workers need more accessible platforms and tasks for screen readers.

- The time taken by visually impaired crowd workers to perform tasks is a crucial problem. This suggests the importance of the development of technologies for reducing the time required to complete tasks.

Related Work

Online Microtasks

Online microtasks are tasks such as categorizing images and drafting articles on crowdsourcing platforms, such as Amazon Mechanical Turk (AMT). Most online tasks are visual and can be run on a personal computer.

Luz et al. provided some types of online tasks including categorization, classification, data verification, data extraction, moderation and tagging of multimedia content, transcription from multimedia content, sentiment analysis and surveys and search relevance.

A few researchers have focused on task design and user interface to enhance the worker performance. Sampath et al. claimed that using cognitively inspired features for task design is a powerful technique to maximize the performance of crowd workers. Rahmanian et al. studied the effects of different user interface designs on the performance of crowdsourcing systems. Manino et al. provided the first theoretical explanation of the accuracy gap between the most popular collection policies.

Calvo et al. evaluated the accessibility of crowdsourcing tasks on Amazon Mechanical Turk and found some problems for visually impaired crowd workers, such as pages that lacked sufficient metadata describing the page language, which could cause a screen reader to have difficulty pronouncing page text.

Visually Impaired Crowd Workers

BSpeak is an accessible crowdsourcing marketplace that enables the visually impaired in developing regions to earn money by transcribing audio files containing speech (Vashistha et al.). Zyskowski et al. presented the first formal study of crowd workers with disabilities via in-depth open-ended interviews of 17 people (disabled crowd workers and job coaches for people with disabilities) and a survey of 631 adults with disabilities. They found several accessibility and usability barriers such as inaccessible task features, inability to find accessible tasks, and time restrictions to complete tasks, limiting the adoption of crowdsourcing marketplaces by blind people.

Research Method

Participants

An online survey was conducted on crowdsourcing platforms for visually impaired and sighted crowd workers named SunnyBank and Crowdworks, which are the main crowdsourcing platforms in Japan. The online survey was approved by the IRB at the Faculty of Library, Information and Media Science at the University of Tsukuba. (approval number: 22-23)

A total of 10 visually impaired and 20 sighted crowd workers were recruited. Each visually impaired crowd worker received compensation of 3,000 Japanese yen for participating in the survey. Each sighted crowd worker received compensation of 500 Japanese yen for participating in the survey. The reward for visually impaired crowd workers was higher because the platform needed more commission to check the visually impaired crowd workers' information, and visually impaired crowd workers were required to answer more questions.

Questions

The survey contained 31 and 26 questions for the visually impaired crowd workers and sighted crowd workers, respectively (See Table 1). The questions were related to the following topics: demographics, visual impairment, use of the crowdsourcing platform, and accessibility problems.

Questions on demographics: They included questions on age, gender, education level, and working conditions.

Table 1: Table of Questions.

Question Number (Q)	**Question**
Q1	Age
Q2	Gender
Q3	If you don't mind, please tell us what kind of visual impairments you have?
Q4	Select letter(s) you can read. (multiple answers is ok.)
Q5	If you select magnified printed letters in Q4, please tell us how you zoom in?
Q6	If you select magnified printed letters in Q4, please tell us what is the preferable number of points?
Q7	Do you use a screen reader?
Q8	If you choose 'Yes' in Q7, please tell us which screen reader you use?
Q9	What is your final education?
Q10	Do you have a regular (daily) job?
Q11	Do you have another job except for crowd-working or a regular job?
Q12	What kind of device are you using when you take part in Amazon Mechanical Turk?
Q13	Which crowdsourcing platform did you use other than Amazon Mechanical Turk?
Q14	Why do you choose these crowdsourcing platforms?

Question Number (Q)	Question
Q15	How did you know the existence of Amazon Mechanical Turk?
Q16	If you choose ’others’ in Q15, please write in detail.
Q17	How did you learn these crowdsourcing platforms?
Q18	If you choose ’others’ in Q17, please write in detail.
Q19	How long did you use a crowdsourcing platform including other platforms (total experience)?
Q20	How often do you finish tasks in a crowdsourcing platform?
Q21	Please write the details about Q20.
Q22	On average, how long do you do crowdsourcing tasks?
Q23	Please write the details about Q22.
Q24	How many tasks did you finish in a crowdsourcing platform including other platforms?
Q25	What kind of tasks did you choose (you could finish)?
Q26	Why do you choose these tasks?
Q27	How did you find the tasks?
Q28	If you write ’searching keywords’ in Q27, please tell us what kind of keywords you searched for?
Q29	What kind of functionality do you expect in a crowdsourcing platform?
Q30	What problems do you have when you are doing the tasks?
Q31	Do you have any suggestions for these problems?

Questions of visual impairment: They include questions on the type of visual impairment, letters they can read, and the use of screen readers. This part is only for visually impaired workers.

Questions of use of the crowdsourcing platform: They included questions on the reasons why they chose a crowdsourcing platform and their experience on crowdsourcing platforms.

Questions of accessibility problems: They included questions on the method of searching for accessibility tasks, the problems that they overcome, a function that they hope for, and suggestions.

Results

Demographics

Age

The visually impaired and sighted participants were aged between 10 and 59 years, with most people between the ages of 40 and 59 for visually impaired crowd workers. There are no participants above the age of 60. Thus, it can be concluded that the visually impaired crowd workers belonged to the "middle age category" (See Table 2).

Table 2. Age

Age range	Visually Impaired Workers	Sighted Workers
10-19 years old	0%	0%
20-29 years old	20%	20%
30-39 years old	10%	40%
40-49 years old	40%	35%
50-59 years old	0%	5%
more than 60 years old	0%	0%

Gender

On SunnyBank, the number of male and female participants was approximately the same. On Crowdworks, there were more male participants than females (See Table 3).

Table 3. Gender

Gender	Visually Impaired Workers	Sighted Workers
Male	50%	60%
Female	50%	40%

Education Level

Table 4 summarizes the education level of the visually impaired and sighted participants. Approximately 70% of the visually impaired participants and 80% of the sighted participants had a university degree at least. Thus, we can conclude that the majority of visually impaired participants were educated, which implies that they possess the ability to perform tasks on crowdsourcing platforms.

Table 4. Education Level

Education Level	Visually Impaired Workers	Sighted Workers
High school	30%	20%
University	50%	75%
Graduated school	10%	5%
Other	0%	0%

Working Conditions

Table 5 details the working conditions of the visually impaired and sighted crowd workers and shows whether visually impaired and sighted crowd workers have regular jobs. Approximately 50% of the visually impaired and 70% of the sighted crowd workers have regular jobs.

Table 5. Working Style

Working Style	Visually Impaired Workers	Sighted Workers
Having regular work	50%	70%
Having no regular work	50%	30%

Visual Impairment

Types of Visual Impairment

All visually impaired participants had a disability certification and all were at level 3 or higher, with vision below 0.07. 80% of the participants were at level 1, with vision below 0.01.

Letters

A total of 80% of the participants could read the braille script, 30% could read magnified printed letters, and 10% could read printed letters. Methods for magnification included lens, magnifying via app or screen, and glasses.

Use of Screen Readers

There were nine screen-reader users. Various screen readers included those by NVDA, PC-Talker, Voice-Over and JAWS.

Use of the Crowdsourcing Platform

Devices

Table 6 indicates that approximately 60% of the visually impaired participants and 95% of the sighted participants used PCs to perform tasks on crowdsourcing platforms.

Table 6. The devices visually impaired and sighted crowd workers use.

Device	Visually Impaired Workers	Sighted Workers
PC	60%	95%
Tablet	10%	5%

Device	Visually Impaired Workers	Sighted Workers
Smartphone	30%	0%

Crowdsourcing Platforms

Furthermore, we investigated other crowdsourcing platforms apart from Sunnybank and Crowdworks. The visually impaired crowd workers also used Crowdworks, Sagoworks and Freelancer. The sighted crowd workers also used Yahoo! Crowdsourcing, Lancers and Zaitakuworks. We discovered there are limited crowdsourcing platforms in Japan. Table 7 summarizes how the participants discovered the crowdsourcing platforms. Table 8 summarizes how the visually impaired and sighted participants learned to use these crowdsourcing platforms.

Table 7. The way visually impaired and sighted crowd workers know the existence of crowdsourcing platforms

Source	Visually Impaired Workers	Sighted Workers
From other people	80%	10%
Searching by yourself	10%	90%
Others	10%	0%

Table 8. The way how visually impaired crowd workers learn these crowdsourcing platforms

Source	Visually Impaired Workers	Sighted Workers
From other people	20%	5%
Searching by yourself	80%	95%

Reason

We investigated the reasons why the participants selected a specific crowdsourcing platform. For visually impaired crowd workers, there were three main criteria: the platform, the reasons included consideration for people with disabilities and simplicity of operation. For the task, the reasons included finding a task that they could do and working with peace of mind even in the current situation, including disabilities and physical conditions. For others, the reason was that more friends used the platform.

For sighted crowd workers, the same three criteria were used: platform, task, and others. For the platform, the reasons included lack of time limit, lots of tasks, matching tasks, major companies and relatively high rewards. For the task, the reasons included the numerous tasks for females. For others, the reasons included being introduced from blogs, YouTube and friends.

Time

Table 9 lists the working hours of the participants per day; 80% of the visually impaired crowd workers worked 1h to 5h per day. 20% of the visually impaired crowd workers worked less than 1h per day. 75% of sighted crowd workers worked 1h to 5h per day. 25% of the sighted crowd workers worked less than 1h per day. We discover that visually impaired crowd workers work longer hours than sighted crowd workers.

Table 9. Working hours of visually impaired and sighted crowd workers per day

Frequency	Visually Impaired Workers	Sighted Workers
Less than 1 hour per day	20%	25%
1-5 hours per day	80%	75%

Frequency

Table 10 lists the number of tasks performed by the crowd workers on the crowdsourcing platforms per month; 60% of visually impaired crowd workers finished 1 to 5 tasks per month and 40% of visually impaired crowd workers finished less than 1 task per month. The reason is that visually impaired workers cannot perform too many tasks. 85% of sighted crowd workers finished more than 10 tasks per month. Six sighted crowd workers finished more than 30 tasks per month. We determine that visually impaired crowd workers finish much fewer tasks than sighted crowd workers.

Table 10: The frequency visually impaired crowd workers finish tasks in a crowdsourcing platform

Frequency	Visually Impaired Workers	Sighted Workers
Less than 1 task per month	40%	0%
1-5 tasks per month	60%	5%
5-10 tasks per month	0%	10%
More than 10 tasks per month	0%	85%

Tasks

For the number of tasks performed by the crowd workers in crowdsourcing platforms including other platforms; 90% of the participants finished less than 20 tasks on the crowdsourcing platforms. These tasks are mainly related to web accessibility tests, surveys, writing, etc.

95% of the sighted crowd workers finished less than 50 tasks in the crowdsourcing platforms. These tasks are mainly related to survey, writing, data input, interview, report of search results, experiments, investigations, etc. We observed that visually impaired crowd workers completed fewer tasks than sighted workers.

Accessibility Problems

The Method of Searching for Accessible Tasks

Methods used by the visually impaired participants for finding tasks included searching, looking for the list on the homepage of crowdsourcing platforms, and recommendations from others. The keywords used by them to search for a task on a crowdsourcing platform included teleworking, disability, simple, beginner, side job, accessibility, writing, questionnaire, point, article and translation. It can be concluded that the participants focused on the type and content of tasks.

Problems

70% of the visually impaired crowd workers mentioned no problems with finishing tasks. However, 40% of the visually impaired participants mentioned almost no suitable tasks for them. Visually impaired crowd workers work longer hours than sighted crowd workers.

Functionalities and Suggestion

The participants desired functionalities such as pausing the time, functionality expansion, screen-reader features, easy understanding of the content, typing error checks, and task recommendation by AI. Other suggestions included more detailed descriptions and simplifying the tasks available.

Discussion and Conclusion

In this study, the survey was conducted in Japanese on crowdsourcing platforms. Most of the Japanese visually impaired crowd workers only performed tasks such as taking surveys, writing articles and accessibility tests. The survey results suggest several ways for using AI to support visually impaired workers. For example, most tasks are simple text-based tasks. More complex text tasks, such as entity-matching tasks (Zhong et al.), and more media-type tasks, such as image tasks (Zhong et al.), can be extended.

Additionally, time is a significant factor to consider. Visually impaired crowd workers spend more time than sighted crowd workers. Reducing the working time is an important issue. The completion time of visually impaired crowd workers can be reduced using several methods.

Zhong et al. suggested that the completion time can be reduced by removing unimportant images from the tasks that use web images and consistency-conscious design of tabular tasks.

The proposals for the future direction are as follows:

- We can support visually impaired workers with AI. For example, most tasks are simple text-based tasks. More complex text tasks, such as entity-matching tasks, and more media-type tasks, such as image tasks, can be included.
- We can extend the time required for task completion for visually impaired crowd workers and provide the estimated completion time.
- We can reduce the completion time of visually impaired crowd workers. For example, completion time can be reduced by removing unimportant images on the task which uses web images and consistency-conscious design of tabular tasks.
- We can change the modality (e.g., visual information to auditory information, text to speech technology, alt text generation) to improve the performance of visually impaired crowd workers, especially screen reader users.

Acknowledgments

This work was partially supported by JSPS KAKENHI Grant Number 22H00508 and 21H03552, Japan. This work was approved by IRB of University of Tsukuba.

Works Cited

Bourne, R., Price, H., Stevens, G. *Global burden of visual impairment and blindness.* Arch Ophthalmol, 2012.

Calvo, R., Kane, S. K., Hurst, A. *Evaluating the accessibility of crowdsourcing tasks on Amazon's mechanical turk.* Proceedings of the 16th international ACM SIGACCESS conference on computers accessibility, pp. 257–258, 2014.

Luz, N., Silva, N., Novais, P. *A survey of task-oriented crowdsourcing* Artificial Intelligence Review, no. 44, pp. 187–213, 2015.

Manino, E., Tran-Thanh, L., Jennings, N. *On the efficiency of data collection for crowdsourced classification.* Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18), pp. 1568-1575, 2018.

Rahmanian, B., Davis, J. G. *User interface design for crowdsourcing systems.* Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces (AVI 2014), pp. 405-408, 2014.

Sampath, H. A., Rajeshuni, R., Indurkhya, B. *Cognitively inspired task design to improve user performance on crowdsourcing platforms.* International Conference for Human-computer Interaction (CHI2014), pp. 3665-3674, 2014.

Vashistha, A., Sethi, P., Anderson, R. *Bspeak: An accessible crowdsourcing marketplace for low-income blind people.* International Conference for Human-computer Interaction (CHI 2018), 2018.

Zhong, Y., Kobayashi, M., Matsubara, M., Morishima, A. *Effects of crowd-in-the-loop alt text addition on the performance of visually impaired workers in online microtasks.* The Transaction of Human Interface Society (2020), vol. 20, pp 251-262, 2020.

Zhong, Y., Kobayashi, M., Matsubara, M., Morishima, A. *Usability of audio matrix microtasks for screen reader users.* The Journal on Technology and Persons with Disabilities (2021), vol. 9, pp 54-70, 2021.

Zhong, Y., Matsubara, M., Kobayashi, M., Morishima, A. *Understandable microtasks with no visual representation.* The Tenth AAAI Conference on Human Computation and Crowdsourcing (HCOMP 2019), 2019.

Zyskowski, K., Morris, M., Bigham, J., Gray, M., Kane, S. *Accessible crowdwork? understanding the value in and challenge of microtask employment for people with disabilities.* Proceedings of the 18th ACM Conference on Computer-Supported Cooperative Work Social Computing (CSCW2015), pp. 1682-1693, 2015.

THE JOURNAL ON TECHNOLOGY AND PERSONS WITH DISABILITIES

Empathy Talk with the Visually Impaired in Design Thinking

Hyung Nam Kim
North Carolina A&T State University
hnkim@ncat.edu

Abstract

Empathy – *a core element in Design Thinking* – enables designers to view the world through users' perspective, leading to successful design practices and outcomes. Yet, engineering students tended to prioritize design solutions to satisfy other stakeholders rather than target users. Such misalignment between theory and practice leads to a research question: how today's engineering students generate and are engaged in empathic conversations with users with visual disabilities in user experience (UX) research. Hence, this study aimed at addressing the knowledge gap by analyzing transcripts of interviews administered by engineering students. In addition, transcripts of interviews by one faculty member were included to be compared with the students. The transcripts were coded by using pre-defined themes, including empathic opportunities, opportunity terminators, and empathic responses (naming, understanding, respecting, supporting, and exploring). This study found that the student researchers had a lack of empathic communication skills. Engineering programs should be adequately amended to train and produce engineers who are equipped with empathy and yield effective design solutions to meet user needs. Future research will, thus, focus on designing educational interventions to properly train students to be competent with empathically connecting with participants in Design Thinking.

Keywords

User-Centered Design, Emotional Intelligence, Designer, Participant, Blind, Visual Impairment.

Introduction

Design Thinking is a systematic design process where a designer is engaged in opportunities to elicit user needs, create designs, evaluate prototypes, gather feedback, and refine designs (Brown and Wyatt; Razzouk and Shute). Empathy – one of the most important elements in Design Thinking – is the ability to view the world through another person's perspective, e.g., to see what users see, to feel what users feel, and to experience things as users experience. Empathy in Design Thinking will help designers to gain a deep understanding of challenges the target user group encounters, discover their latent needs, and create solutions that accommodate adequately the target user group (Kouprie and Visser; Smeenk, Sturm and Eggen). Empathy is also necessary for designers working with users with visual disabilities (Kim).

Education of empathy has traditionally been focused on children and adolescent development, but there is a growing attention to students in post-secondary education systems (Hess and Fila; von Unold et al.). For example, Bairaktarova et al. introduced a conceptual framework to integrate empathic design thinking in the engineering curriculum and also empirically conducted empathy-based design practices with over 100 engineering students. They found that the students were able to better understand user needs, resulting in more feasible design solutions.

Yet, there is evidence that engineering students are less likely to conduct user experience research with empathy. Congruence between *belief* of importance of empathic designs and *behavior* of applying empathic design approaches is necessary. However, Guanes et al. observed that engineering students often failed to prioritize design solutions for the target user group although they were aware of importance of applying empathic design approaches. Such misalignment leads us to the research question: *how today's engineering students generate and* *share empathic conversations with users while conducting user experience research.*

This study aims to advance understanding of the degree to which engineering students have empathic conversations with users in user experience research. A faculty member who has extensive experience with Design Thinking is also invited to be compared with engineering students.

Methods

This study used the transcripts of interviews that were administered by an undergraduate research assistant, a graduate research assistant, and a faculty member. The undergraduate research assistant was a senior student who took human factors courses and was previously engaged in a human factors research project that aimed to develop a multimedia-based user guide for smartphone users with visual disabilities. The graduate research assistant was a second-year doctoral student who took human factors courses and was previously engaged in several human factors research projects for people with visual disabilities. The faculty member was engaged in multiple research projects over 16 years, aiming to improve the quality of life for people with various visual acuity levels (ranging from normal vision to blindness). The three researchers completed an educational program for the protection of human subjects in research via the Collaborative Institutional Training Initiative (CITI).

This study was based on a *N-of-1* trial approach to produce a large set of data. In 2019, the faculty member conducted interviews with 20 participants (a total of 200 hours). In 2020, the undergraduate and graduate research assistants also did so with seven participants (a total of 70 hours) and 23 participants (a total of 230 hours), respectively. The participants met inclusion criteria: visual impairment/blindness (visual acuity equal to or poor than 20/70), no other sensory impairments, and community-dwelling adults (age 18 or older). The participants were instructed to share their experiences and challenges associated with vision loss. A semi-structured interview was employed in which a short list of guiding questions were supplemented by follow-up and probing questions, depending on the participants' responses. For example, they were asked to think about how they became visually impaired/blind and how they have performed daily activities at home, workplace, and/or community, which was followed by the question "*What functional challenges have you encountered associated with vision loss?*" They were asked to think about how they were engaged in social interactions with family members, friends, and/or colleagues, which was followed by the question "*What social challenges have you encountered associated with vision loss?*" At that time, all the researchers were not aware that the interview transcripts would be used for this study. Hence, their behaviors were reasonably assumed to be natural.

The transcripts were coded by referring to the codebook of Suchman et al. and the code book of Back et al., The codebook of Suchman et al., helped to code empathic interactions (i.e., empathic *opportunities*, empathic *responses*, and empathic *opportunity terminators*). The empathic opportunity terminator refers to a researcher's statement that immediately follows an empathic opportunity but directs the interview away from the stated emotion. The codebook of Back et al. helped to code "empathic responses" in more detail, by breaking them into *naming, understanding, respecting, supporting,* and *exploring*.

Results

Participants' Expressions of Emotions and Researchers' Responses

Table 1 shows the frequency of cases where *Empathic Opportunity, Empathic Response, and Empathic Opportunity Terminator* occurred. Between- and within-subject data analyses were also presented in the following sections.

Table 1. Frequency of cases on average for empathic opportunities, responses, and terminators.

Participant type	**Empathic opportunity** (Mean ± SD)	**Empathic response** (Mean ± SD)	**Empathic opportunity terminator** (Mean ± SD)
Undergraduate student	5.43 ± 1.51	2.00 ± 1.53	4.86 ± 1.68
Graduate student	5.65 ± 1.67	4.52 ± 2.61	4.52 ± 0.99
Faculty	7.37 ± 3.88	5.95 ± 4.14	2.68 ± 2.60

Between-subject Data Analysis

Empathic Response. A Kruskal-Wallis test found a significant difference in the frequency of Empathic Response between the three researchers, $H(2) = 6.90$, $p = 0.03$. Mann-Whitney tests were used to follow up on the significant finding. The frequency of Empathic Response of the undergraduate student was less than that of the graduate student and the faculty, $U = 31.50$, $r = -0.44$ and $U = 29.5$, $r = -0.42$, respectively.

Empathic Opportunity Terminator. A Kruskal-Wallis test found a significant difference in the frequency of Empathic Opportunity Terminator between the three researchers, $H(2) = 13.57$, $p < 0.01$. Mann-Whitney tests showed that the frequency of Empathic Opportunity Terminator of the faculty member was less than that of both undergraduate and graduate students, $U = 24$, $r = -0.49$ and $U = 85$, $r = -0.53$, respectively.

Within-subject Data Analysis

Undergraduate student. A Friedman's test found a significant difference in the undergraduate student's responses, $\chi^2(2) = 6.09$, $p < 0.05$. Wilcoxon signed-rank tests were used to follow up on the significant finding. The frequency of Empathic Response was less than that of Empathic Opportunity, $z = -2.21$, $p < 0.05$, while the frequency of Empathic Opportunity Terminator was not significantly different from that of Empathic Opportunity, $z = -1.34$, $p = 0.18$.

Graduate student. A Friedman's test found a significant difference in the graduate student's responses, $\chi^2(2) = 9.46$, $p < 0.01$. Wilcoxon signed-rank tests showed that the frequency of Empathic Response was less than that of Empathic Opportunity, $z = -2.14$, $p < 0.05$, but the frequency of Empathic Opportunity Terminator was less than that of Empathic Opportunity, $z = -2.96$, $p = 0.03$.

Faculty. A Friedman's test found a significant difference in the faculty member's responses, $\chi^2(2) = 21.17$, $p = 0.00$. Wilcoxon signed-rank tests showed that the frequency of Empathic Response was less than that of Empathic Opportunity, $z = -2.53$, $p = 0.01$; the frequency of Empathic Opportunity was greater than that of Empathic Opportunity Terminator, $z = -3.38$, $p < 0.01$; and the frequency of Empathic Response was greater than that of Empathic Opportunity Terminator, $z = -2.80$, $p < 0.01$.

Researchers' Empathic Responses in More Detail

The three researchers' empathic responses were broken down into exploring, understanding, naming, respecting, and supporting (see Figure 1). The dominant type of empathic response was "supporting" in the undergraduate student, "understanding" in the graduate student, and "respecting" in the faculty member. Between- and within-subject data analyses were presented in the following sections.

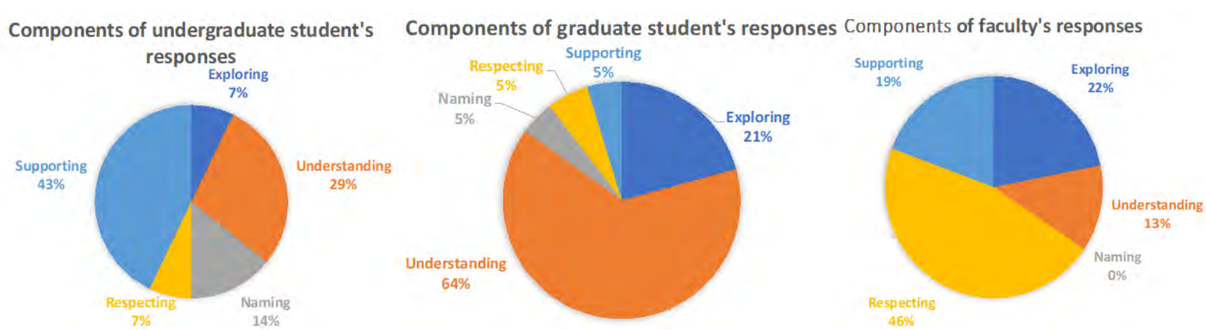


Fig. 1. Different types of empathic responses between the three researchers.

Between-subject Data Analysis

Understanding. A Kruskal-Wallis test found a significant difference in the frequency of *understanding* between the three researchers, $H(2) = 19.78$, $p < 0.01$. Mann-Whitney tests showed that the frequency of *understanding* of the graduate student was greater than that of the undergraduate student, $U = 17.50$, $r = -0.57$ and that of the faculty member, $U = 66$, $r = -0.61$.

Respecting. A Kruskal-Wallis test found a significant difference in the frequency of *respecting* between the three researchers, $H(2) = 18.08$, $p < 0.01$. Mann-Whitney tests showed that the frequency of *respecting* of the faculty member was greater than that of the undergraduate and graduate students, $U = 21.50$, $r = -0.53$ and $U = 80$, $r = -0.59$, respectively.

Supporting. A Kruskal-Wallis test found a significant difference in the frequency of *supporting* between the three researchers, $H(2) = 7.52$, $p = 0.02$. Mann-Whitney tests showed that the frequency of *supporting* of the graduate student was less than that of the undergraduate student and the faculty, $U = 48$, $r = -0.38$ and $U = 133.50$, $r = -0.40$, respectively.

Within-subject Data Analysis

Undergraduate student. A Friedman's test found no significant difference in the undergraduate student's empathic responses such as *exploring, understanding, naming, respecting,* and *supporting*.

Graduate student. A Friedman's test found a significant difference in the graduate student's empathic responses, $\chi^2(4) = 52.01$, $p < 0.01$. Wilcoxon signed-rank tests showed that the frequency of *understanding* was greater than that of the other response types, i.e., *exploring* ($z = -3.60$, $p < 0.01$), *naming* ($z = -3.94$, $p < 0.01$), *respecting* ($z = -3.95$, $p < 0.01$), and *supporting* ($z = -4.13$, $p < 0.01$). In addition, the frequency of *exploring* was greater than that of *naming* ($z = -2.42$, $p = 0.02$), *respecting* ($z = -2.35$, $p = 0.02$), and *supporting* ($z = -2.44$, $p = 0.02$).

Faculty. A Friedman's test found a significant difference in the faculty member's empathic responses, $\chi^2(4) = 23.15$, $p < 0.01$. Wilcoxon signed-rank tests showed that the frequency of *naming* was less than that of all the other response types, i.e., *exploring* ($z = -2.95$, $p < 0.01$), *understanding* ($z = -2.41$, $p = 0.02$), *respecting* ($z = -3.31$, $p < 0.01$), and *supporting* ($z = -2.84$, $p < 0.01$). Another significance was that the frequency of *respecting* was greater than that of *understanding* ($z = -2.59$, $p < 0.01$) and *supporting* ($z = -2.50$, $p = 0.01$).

Discussion

Participants' Expressions of Emotions and Researchers' Responses

The participants made emotional comments (i.e., empathic opportunities) while interacting with the researchers (the undergraduate student, the graduate student, and the faculty). However, all the three researchers tended to miss many opportunities of responding to participants' emotional comments. The results suggest that there has been a lack of adequate training for the researchers to develop empathic communication skills.

Several efforts (Bairaktarova et al.; von Unold et al.) have been made to incorporate Design Thinking in curriculums for college students; however, many students still tend to encounter a difficulty empathizing with users (Smeenk, Sturm, Terken, et al.). Shambaugh and Beacham argued that there have been limitations of teaching Design Thinking; for example, Design Thinking was considered as a subjective idea in many design-related curriculums and not considered as an explicit learning outcome. Hence, Design Thinking often occurred as an incidental learning outcome, instead of a well-planned learning outcome (Shambaugh and Beacham). Afroogh et al. also pointed out that creating and sustaining an inclusive and effective community resilience approach requires empathy, which was often disregarded in the existing engineering education and practice. Smeenk, Sturm and Eggen argued that a lack of empathy tends to occur under various circumstances, e.g., when designers fail to meet, collaborate, and/or connect with users in person; when designers struggle with insufficient time or budget; and when designers have a lack of ability or willingness to obtain empathy. Given the results, this study recommends that existing engineering programs should adequately be amended to produce engineers who can yield effective solutions to meet user needs while being emotionally connected with users, which has also been voiced by other studies (Evans; Razzouk and Shute).

Researchers' Empathic Responses in Detail

The response of *understanding* was significantly observed in the graduate student as compared to the undergraduate student and the faculty. The result suggests that the graduate student had a stronger tendency to confirm and summarize what he was hearing while conducting interviews. On the contrary, the response of *supporting* was significantly missing in the graduate student as compared to other researchers. The results suggest that the graduate student might have focused more on "accurately" and "objectively" gathering data; thus, the graduate student made less efforts to "subjectively" making emotional connection with participants. Emotional connection between an interviewer and an interviewee contributed to a rich interview experience, resulting in a deep comfort with one another (Ross). Yet, Thomas and McDonagh pointed out that it would be challenging to design "with" users instead of "for" users. Despite the challenge, interactive co-design practices must be secured to produce design outcomes that accommodate user needs. Such design practices could be facilitated via shared language, collaboration, and empathy (Thomas and McDonagh).

The response of *respecting* was significantly observed in the faculty as compared to the undergraduate and graduate students. The faculty showed a stronger tendency to pay attention to and praise participants' coping strategies for emotional struggles. It would help the faculty to develop a deep emotional connection with participants. The importance of building strong connection with participants is well documented in the literature (Anderson et al.). For example, Råheim et al. conducted focus groups to explore the experiences of health researchers who conducted qualitative research. The health researchers considered that it was essential to listen to participants' stories, show respect, and gain trust from the participants because the effectiveness of their qualitative research was likely to be influenced by participants' willingness to share their experiences and thoughts about topics in question. Bay-Cheng argued that it is important for researchers to make an opportunity to develop positive, non-judgmental relationships with participants during interviews as participants' comfort increases, participants are more likely to share details of their experience. Bay-Cheng also argued that such positive, non-judgmental relationships are applied to both qualitative and quantitative research because researchers interact with participants during recruitment, informed consent, and debriefing. This study suggests that researchers should be well trained to emotionally connect with participants throughout the entire research processes by pursing mutual respect and shared goals in design.

Conclusion

In Design Thinking, empathy is critical for designers to deeply understand users' needs for whom they are designing. Several tools and techniques have been introduced to facilitate empathic design processes, aiming to support designers to "*step into the user's shoes*" and "*walk the user's walk*" in order to produce products that meet the user needs (Ghajargar et al.). However, this study found evidence that engineering student researchers are less likely to empathically connect with participants due to a lack of empathic communication skills. Future research with a larger sample of researchers (undergraduate, graduate, and faculty) will focus on developing educational interventions to properly train students to be competent with empathically connecting with participants.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 1831969.

Works Cited

Afroogh, Saleh et al. “Empathic Design in Engineering Education and Practice: An Approach for Achieving Inclusive and Effective Community Resilience.” *Sustainability*, vol. 13, no. 7, 2021, p. 4060.

Anderson, Nicholas et al. “Participant-Centric Initiatives: Tools to Facilitate Engagement in Research.” *Applied & translational genomics*, vol. 1, 2012, pp. 25-29.

Back, Anthony L et al. “Approaching Difficult Communication Tasks in Oncology 1.” *CA: a cancer journal for clinicians*, vol. 55, no. 3, 2005, pp. 164-77.

Bairaktarova, Diana et al. “Beyond Surface Knowledge: An Exploration of How Empathic Design Techniques Enhances Engineers Understanding of Users’ Needs.” *Int. J. Eng. Educ*, vol. 32, no. 1, 2016, pp. 111-22.

Bay-Cheng, Laina Y. “Beyond Trickle-Down Benefits to Research Participants.” *Social Work Research*, vol. 33, no. 4, 2009, pp. 243-47.

Brown, Tim and Jocelyn Wyatt. “Design Thinking for Social Innovation.” *Development Outreach*, vol. 12, no. 1, 2010, pp. 29-43.

Evans, Donovan L. “Design in Engineering Education: Past Views of Future Directions.” *Engineering Education*, vol. 80, no. 5, 1990, pp. 517-22.

Ghajargar, Maliheh et al. “Empathy Workshop: When Project Team and Pilot Users Exchange Experiences.” *The Design Journal*, vol. 20, no. sup1, 2017, pp. S3837-S48.

Guanes, Giselle et al. “Empathic Approaches in Engineering Capstone Design Projects: Student Beliefs and Reported Behaviour.” *European Journal of Engineering Education*, vol. 47, no. 3, 2022, pp. 429-45.

Hess, Justin L and Nicholas D Fila. “The Manifestation of Empathy within Design: Findings from a Service-Learning Course.” *CoDesign*, vol. 12, no. 1-2, 2016, pp. 93-111.

Kim, Hyung Nam. “Guidance for Best Practices for Research Involving Participants with Visual Disabilities.” *ergonomics in design*, 2021, pp. 1-7, doi:10.1177/10648046211058248.

Kouprie, Merlijn and Froukje Sleeswijk Visser. “A Framework for Empathy in Design: Stepping into and out of the User's Life.” *Journal of Engineering Design*, vol. 20, no. 5, 2009, pp. 437-48.

Råheim, Målfrid et al. “Researcher–Researched Relationship in Qualitative Research: Shifts in Positions and Researcher Vulnerability.” *International journal of qualitative studies on health and well-being*, vol. 11, no. 1, 2016, p. 30996.

Razzouk, Rim and Valerie Shute. “What Is Design Thinking and Why Is It Important?” *Review of Educational Research*, vol. 82, no. 3, 2012, pp. 330-48.

Ross, Lori E. “An Account from the Inside: Examining the Emotional Impact of Qualitative Research through the Lens of “Insider” Research.” *Qualitative Psychology*, vol. 4, no. 3, 2017, p. 326.

Shambaugh, Neal and Cindy Beacham. “Starting a Dialogue on Design Thinking.” *Critically Engaging the Digital Learner in Visual Worlds and Virtual Environments*, edited by Maria Avgerinou et al., The International Visual Literacy Association 2010.

Smeenk, Wina et al. “A Comparison of Existing Frameworks Leading to an Empathic Formation Compass for Co-Design.” *International Journal of Design*, vol. 13, no. 3, 2019, http://www.ijdesign.org/index.php/IJDesign/article/view/3406.

Smeenk, Wina et al. “A Systematic Validation of the Empathic Handover Approach Guided by Five Factors That Foster Empathy in Design.” *CoDesign*, vol. 15, no. 4, 2019, pp. 308-28.

Suchman, Anthony L et al. “A Model of Empathic Communication in the Medical Interview.” *JAMA*, vol. 277, no. 8, 1997, pp. 678-82.

Thomas, Joyce and Deana McDonagh. “Empathic Design: Research Strategies.” *The Australasian medical journal*, vol. 6, no. 1, 2013, pp. 1-6.

von Unold, Benedikt et al. “Implications of Contextual Empathic Design for Engineering Education.” *2018 ASEE Annual Conference & Exposition*, 2018. doi:10.18260/1-2--30626.

Various Sociodemographic Factors and User Privacy Awareness

Hyung Nam Kim
North Carolina A&T State University
hnkim@ncat.edu

Abstract

As technology advances, many people with visual disabilities take advantage of emerging technologies (e.g., Seeing AI, TapTapSee, and BeMyEyes apps) to gain better access to information. Yet, there is a likelihood that those with visual disabilities are vulnerable to privacy threats as they are less likely to adequately read visual cues and software notifications that are designed to help users to protect themselves and their data from privacy invasion. Unfortunately, little is known about how they perceive privacy policies of the assistive apps, especially depending on sociodemographic backgrounds. To address the knowledge gap, this study invited 30 individuals with visual disabilities to measure their understanding level of privacy policies of the apps but also their tendency level to adopt the apps after being educated on the privacy policies. Individual differences were found, depending on their sociodemographic backgrounds (i.e., vision, age, education, health conditions, race/ethnicity, and living conditions). Given the results, this study suggests that educational interventions are needed to properly inform users with visual disabilities about how their personal information are used, shared, and protected while using the assistive apps, which is anticipated to contribute to safe practices of use.

Keywords

Digital Privacy, Visual Impairment, Blindness, Independent Living, Assistive Apps.

Introduction

Today, a great number of people with visual disabilities take advantage of mobile assistive technologies that enable users to obtain better access to information through a mobile phone camera. Seeing AI, TapTapSee, and BeMyEyes apps are the mostly commonly used camera-based assistive technology apps among people with visual disabilities (Dockery and Krzystolik). For example, Microsoft’s Seeing AI app is equipped with artificial intelligence technology to enhance the capability of reading texts, identifying people/objects, and audibly describing them to users who have visual disabilities. The TapTapSee app takes advantage of an Image Recognition Application Programming Interface (API) in recognizing images. In contrast to the TapTapSee and Seeing AI apps, the BeMyEyes app runs with supports from online volunteers who read and recognize various information in real time when people with visual disabilities share information via a mobile phone camera.

Yet, users may reveal their personal information (e.g., credit cards, bank statements, medical bills, and so on) while using the assistive apps. That personal information would then be shared with the app companies, their third party, and/or other users online. It is reasonable to argue that users with visual disabilities are more vulnerable to being exposed to privacy threats and privacy invasion as they are less likely to fully benefit visual cues and software programs that are designed to help users to notice potential cybersecurity threats (Inan et al.). As we are living in a data-driven world, data breaches could significantly affect a great number of people. Cybersecurity concerns are on the rise around the world. Over 2,000 Internet crime complaints were reported per day in 2020 (Federal Bureau of Investigation). The well-known digital security company, Norton pointed out that more than half of all customers experienced a cybercrime, i.e., nearly one in three falling victim in 2020 alone.

However, little is known about how users with visual disabilities perceive the privacy policies of the camera-based assistive apps as well as whether they are more (or less) likely to adopt the apps after being educated on the privacy policies, especially depending on users' sociodemographic backgrounds. To address the knowledge gap, this study conducted interviews with those apps' users with visual disabilities.

Methods

Participants

Inclusion criteria were English speaking, 18 years old or older, visual acuity equal to or poor than 20/70, and user experience with Seeing AI, BeMyEyes, and TapTapSee apps (named "experienced users"). This study also included "prospective users" who have not used the apps but are interested in using the apps in the future. A convenience sample of 30 individuals with visual disabilities were invited, accounting for 24 experienced users and 6 prospective users (see Table 1).

Table 1a. Characteristics of the Participants – Vision Loss.

Vision Loss (visual acuity level)	**Experienced Users (n = 24)**	**Prospective Users (n = 6)**
Visual Impairment (from 20/200 to 20/1200)	7	3
Blind (from 20/1200 to no light perception)	17	3
Duration of vision loss (years)	29.17 ± 21.46	35.17 ± 22.75

Table 1b. Characteristics of the Participants – Onset of Vision Loss.

Onset of Vision Loss (years)* *Participants with early-onset had lost their vision before 11 years of age (Voss et al.).	**Experienced Users (n = 24)**	**Prospective Users (n = 6)**
Early onset	n = 7 (2.14 ± 3.08)	n = 2 (0.00 ± 0.00)
Late onset	n = 17 (39.29 ± 20.63)	n = 4 (47.00 ± 29.54)
Age (years)	57.63 ± 18.43	66.50 ± 15.08

Table 1c. Characteristics of the Participants – Gender.

Gender	Experienced Users (n = 24)	Prospective Users (n = 6)
Male	9	4
Female	15	2

Table 1d. Characteristics of the Participants – Race/Ethnicity.

Race/Ethnicity	Experienced Users (n = 24)	Prospective Users (n = 6)
African American	9	4
European American	14	1
Others	1	1

Table 1e. Characteristics of the Participants – Head of Household.

Head of Household	Experienced Users (n = 24)	Prospective Users (n = 6)
Living alone	10	4
With family, relatives, or friends	14	2

Table 1f. Characteristics of the Participants – Education.

Education	Experienced Users (n = 24)	Prospective Users (n = 6)
High school or equivalent	10	2
Associate	4	3
Bachelors	4	1
Masters	6	0

Table 1g. Characteristics of the Participants – Health Status.

Health Status	Experienced Users (n = 24)	Prospective Users (n = 6)
Chronic illness	8	1
Healthy	16	5

Materials

A quiz was employed to assess the degree to which users are knowledgeable about user privacy policies. The quiz consisted of 42 true-or-false statements that were extracted from the homepages of the assistive technology apps' companies. The selected policies covered comprehensively various aspects, such as "information the company collects", "sharing of the collected information", "storing of the collected information", and "user capabilities to access and control their personal data." Users' tendency to adopt the apps was assessed using the inquiries developed by Gao et al., yet this study chose particularly seven inquiries relevant to trust in privacy and security.

Procedures

Participants were individually invited to interviews (~60 minutes) by phone. The quiz was administered with experienced users to measure the degree to which they were aware and understood the user privacy policies of the apps they have used. An interviewer read out loud each quiz question, and participants responded verbally by indicating either true or false. Participants were allowed to ask the interviewer to repeat the quiz questions as needed. There was no time limit, such that participants could spend as much time as they need. After completing the quiz session, participants were informed of correct answers. An interviewer provided additional explanations as needed, which was an opportunity for participants to be educated about user privacy policies of the apps.

The adoption inquiries were administered twice – before and after quiz – to assess the degree to which the quiz-based education changes participants' tendency to adopt the apps. In addition to experienced users, prospective users were instructed to participate in the same quiz-based education session as we were interested in exploring the degree to which they change their tendency to adopt the apps after being educated. Non-parametric statistics (e.g., Mann-Whitney test and Spearman's rho correlation) were used to analyze the data as they were not normally distributed.

Results

This study found individual differences associated with the quiz scores and the tendency to adopt the assistive apps, depending on sociodemographic backgrounds (e.g., vision conditions, age, educational backgrounds, health conditions, race/ethnicity, and living conditions).

Quiz for User Privacy Policies

Quiz scores of experienced users were analyzed with a Mann-Whitney Test. The quiz score of participants with late onset of vision loss (i.e., one who lost vision after 11 years of age, (Voss et al.)) was higher (0.85 ± 0.15) than that of those with early onset (0.64 ± 0.28), $U = 27.5$, $r = -0.42$. Spearman's rho correlation showed that the duration (years) of vision loss was negatively correlated with the quiz score, $p = 0.002$, $r = -0.61$.

Tendency to Adopt the Assistive Apps

Both experienced and prospective user groups showed significant differences in adoption tendency, depending on age, education, and health conditions.

Experienced Users

Age. A Mann-Whitney Test showed that the tendency to adopt the assistive apps was greater in younger participants (6.86 ± 0.31) than older participants (6.68 ± 0.26) after the quiz-based education, $U = 34.50$, $r = -0.44$.

Educational Background. A Mann-Whitney Test showed that the tendency to adopt the assistive apps was greater in participants without higher education (6.92 ± 0.15) than participants with higher education (6.62 ± 0.37) after the quiz-based education, $U = 33.00$, $r = -0.48$.

Health Conditions. A Mann-Whitney Test showed that the tendency to adopt the assistive apps was greater in participants with health conditions (6.96 ± 0.10) than healthy participants (6.71 ± 0.33) after the quiz-based education, $U = 30.00$, $r = -0.44$.

Prospective Users

As shown in Figure 1, a descriptive statistic helped to compare the data of prospective users by demographic backgrounds. After the quiz-based education, the greater tendency to adopt the assistive apps was found in the following prospective user groups: older participants, participants living with others, European American participants, and participants with blindness.

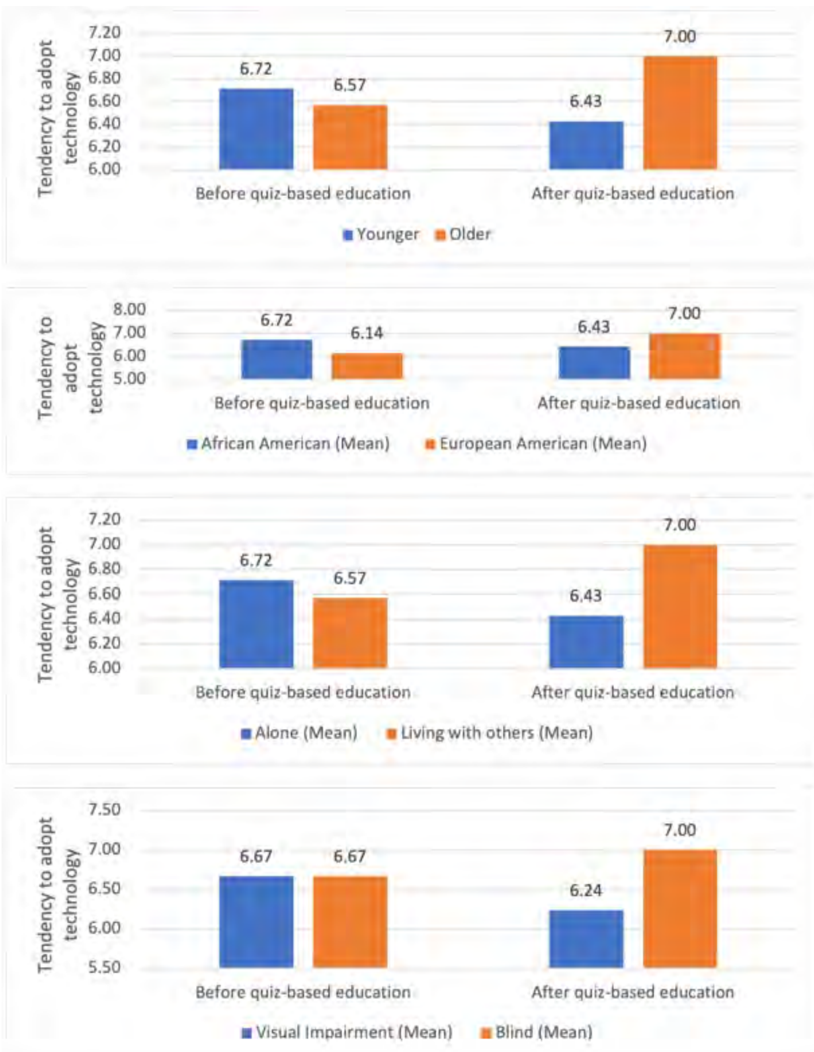


Fig. 1. Mean values of tendency to adopt the assistive apps among prospective users.

Discussions

Quiz for User Privacy Policies

Experienced users with shorter duration of and late onset of vision loss were those who lived with sight for a long period of time in early life. Thus, they did not have any limitation in terms of reading texts and seeing images associated with user privacy policies as well as having actual hands-on experiences with a variety of technologies before losing vision. They might have a number of opportunities to gain deep understanding of user privacy policies about various technologies. Such prior learning might have helped them to understand the user privacy concepts and policies of the assistive technology apps. On the other hand, it also infers that their counterparts who have lost their vision in early life and in the recent past might struggle with comprehensively understanding user privacy policies. There is an urgent need to offer them an adequate means to be well informed of user privacy policies. Thus, they could better protect themselves from security threats and privacy risks.

Tendency to Adopt the Assistive Apps

Experienced Users

Age. Younger participants showed stronger tendency to adopt the apps after the quiz-based education. It may suggest that user privacy education is more effective to younger people with visual disabilities, leading to greater trust on the assistive apps, while older people with visual disabilities might have been more cautious to protect them from being exposed to privacy risks. The age-related individual differences in online privacy awareness are well documented in the literature (Bellman et al.; Mittal and Ilavarasan; Zeissig et al.). For example, Cho et al. conducted a study with 1261 Internet users and found older individuals were more concerned about privacy issues as compared to their younger peers.

Educational Background. Participants without higher education were more likely to adopt the assistive apps. The result infers that participants with higher education might have been more likely to be equipped with conservative attitudes toward using the assistive apps. They might have had more opportunities to be informed of privacy breaches via news media or personal research, probably leading to lower trust on privacy policies. Similar results are also found in the literature (Ojala Burman). Cho et al. also found that Internet users with a higher educational background were more concerned about privacy issues as compared to their counterparts.

Health Conditions. Participants with health conditions were more likely to adopt the assistive apps after the quiz-based education. They might have perceived that the apps would be safe to use and/or willing to compromise their privacy in order to have access to information that were not accessible to them previously. That information might be beneficial to their health conditions and self-care practices, probably leading to more motivation to keep using the assistive apps. Patients’ willingness to accept the privacy breach to obtain access to health information was well documented in the literature (Hale and Kvedar).

Besides technology adoption, the results also provide important insights into teaching and learning in user privacy and security domains. For example, the results infer that the quiz-based education works for those who are younger, without higher education degrees, and with health issues. On the other hand, the results also infer that there is a need to redesign the way of teaching their counterparts. Future research will, thus, refine the education contents to address unresolved privacy-relevant concerns of those counterparts. As this study did not consider individual differences in learning styles, e.g., active/reflective, visual/verbal, sensing/intuitive, and sequential/global (Felder and Silverman), future research will also find adequate education delivery methods to accommodate different learning types of people with visual disabilities.

Prospective Users

Among the prospective users, those who were blind, older, living with others, and European Americans showed greater tendency to adopt the assistive apps. The results infer that the current quiz-based education is an effective teaching and learning style suitable to those prospective user groups while that style is less effective to their counterpart groups. There is a need to refine the educational contents and teaching method that should work for those counterpart groups. By doing so, more people with various sociodemographic backgrounds could feel safe to use the assistive apps, ultimately contributing to technology adoption and enhancement of quality of life and independence.

Conclusions

This study found evidence that, depending on sociodemographic backgrounds, certain users with visual disabilities had a lower level of understanding of privacy policies of mobile camera-based assistive apps as well as a lower level of tendency to adopt the apps. Given the results, this study recommends that educational interventions be adequately designed to teach users with visual disabilities how to safely use the apps to keep personal information secure.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 1831969.

Works Cited

Bellman, Steven et al. “International Differences in Information Privacy Concerns: A Global Survey of Consumers.” *The Information Society*, vol. 20, no. 5, 2004, pp. 313-24.

Cho, Hichang et al. “A Multinational Study on Online Privacy: Global Concerns and Local Responses.” *New Media & Society*, vol. 11, no. 3, 2009, pp. 395-416.

Dockery, Dominique and Magdalena Krzystolik. “The Use of Mobile Applications as Low-Vision Aids: A Pilot Study.” *Rhode Island Medical Journal*, vol. 103, no. 8, 2020, pp. 69-72.

Federal Bureau of Investigation. “Internet Crime Report 2020.” https://www.ic3.gov/Media/PDF/AnnualReport/2020_IC3Report.pdf. Accessed September 23 2022.

Felder, Richard M and Linda K Silverman. “Learning and Teaching Styles in Engineering Education.” *Engineering Education*, vol. 78, no. 7, 1988, pp. 674-81.

Gao, Shang et al. “Developing an Instrument to Measure the Adoption of Mobile Services.” *Mobile Information Systems*, vol. 7, no. 1, 2011, pp. 45-67.

Hale, Timothy M and Joseph C Kvedar. “Privacy and Security Concerns in Telehealth.” *AMA Journal of Ethics*, vol. 16, no. 12, 2014, pp. 981-85.

Inan, Fethi A et al. “Internet Use and Cybersecurity Concerns of Individuals with Visual Impairments.” *Journal of Educational Technology & Society*, vol. 19, no. 1, 2016, pp. 28-40.

Mittal, Shweta and P Vigneswara Ilavarasan. “Demographic Factors in Cyber Security: An Empirical Study.” *Conference on e-Business, e-Services and e-Society*, Springer, 2019, pp. 667-76.

Norton. “2021 Norton Cyber Safety Insights Report.” https://now.symassets.com/content/dam/norton/campaign/NortonReport/2021/2021_NortonLifeLock_Cyber_Safety_Insights_Report_Global_Results.pdf. Accessed September 2022.

Ojala Burman, Emma. “Impact of Demographic Factors on Information Security Awareness: A Study on Professionals and Students in Sweden.” *School of Informatics*, vol. Master’s, University of Skövde, 2021. http://www.diva-portal.org/smash/get/diva2:1576910/FULLTEXT01.pdf.

Voss, Patrice et al. “Early-and Late-Onset Blind Individuals Show Supra-Normal Auditory Abilities in Far-Space.” *Curr Biol*, vol. 14, no. 19, 2004, pp. 1734-38.

Zeissig, Eva-Maria et al. “Online Privacy Perceptions of Older Adults.” *International Conference on Human Aspects of IT for the Aged Population*, Springer, 2017, pp. 181-200.

THE JOURNAL ON TECHNOLOGY AND PERSONS WITH DISABILITIES

Investigating Accessibility Issues in Scheduling Coordination for Visually Impaired Computer Users

Takahiro Miura[1], Hiroki Watanabe[2], Masaki Matsuo[2], Masatsugu Sakajiri[2], Junji Onishi[2]
National Institute of Advanced Industrial Science and Technology, Japan[1]
Tsukuba University of Technology, Japan[2]
miura-t@aist.go.jp, k182311@cc.k.tsukuba-tech.ac.jp, matsuo@cs.k.tsukuba-tech.ac.jp, sakajiri@g.tsukuba-tech.ac.jp, ohnishi@g.tsukuba-tech.ac.jp

Abstract

Although various scheduling methods have merits and demerits, accessibility issues, particularly for visually impaired users, have not been adequately clarified. Herein, we clarified accessibility issues with scheduling interfaces for visually impaired people and propose measures to deal with them. We first created four different types of modeled-scheduling interfaces. Then, we asked the visually impaired people who used a speech-to-text to register for a simulated schedule with these interfaces and answer questionnaires about their usability. The results revealed that the form with radio buttons for circle (○, available for participation), triangle (△, partially available), and cross mark (×, unavailable) and with a comment field for each candidate date received the highest usability with a significant score.

Keywords

Visually impaired people, web-based scheduling coordination, usability, radio buttons, comment fields, Big Five.

Introduction

In our daily lives, we have a need to manage and coordinate our schedules, and there are many methods available for adjusting them in various situations; some situations require adjusting one's schedule, while others involve coordinating schedules among multiple people to maximize the number of participants. The specific methods for schedule coordination include e-mail, chat tools, and web tools. Although each method has its merits and demerits, accessibility issues, particularly for visually impaired users, have not been sufficiently clarified. Previous studies have proposed systems to reduce the time and effort required for schedule coordination (Motomura et al. 3126). Some studies examined the usability of calendar interfaces qualitatively (Carvalho et al. 2029) and interface design based on participatory design approaches (Ghidini et al. 5700); however, few quantitative research has been conducted on the ease of input by the type of interface elements and choices, as well as on how to interact with people with various traits and determine their availability for schedule coordination.

Therefore, herein, we clarify the accessibility and usability issues of each scheduling method for visually impaired users and propose measures to deal with them. After clarifying these issues using questionnaires, we propose a system that improves the efficiency and convenience of scheduling based on the acquired information. Our research questions are as follows:

Q1. What are the requirements that a scheduling application should satisfy to be easy in use for the visually impaired?

Q2. What are the preferences of the visually impaired for a scheduling application? Which type of people like/dislike this type of application?

Method

Visually impaired people who used a speech-to-text tool were asked to register a simulated schedule with the scheduling interfaces. Then, they were asked to answer questionnaires about their personalities and the usability of these interfaces. Finally, after studying the results, we analyzed the elements that visually impaired people consider when registering and planning their schedules.

Participants

This study includes fourteen persons, twelve men and two women (Mean age: 26.5). Eleven and three persons are with and without visual impairments, respectively. Eight of the visually impaired ones are congenitally blind and had a disability certificate issued in Japan. Nine and two of the visually impaired had low vision and total blindness, respectively. One of the total blinds has no light perception, while the other has motus manus. All the visually impaired participants could use their computers with screen readers; whereas, the others could not. Two of the low-vision participants has never used web-based scheduling applications; whereas, the others have some experience with them.

Evaluated Scheduling Interface

We created four types of modeled schedule-registration forms based on a preliminary analysis of existing web-based scheduling applications. Details of these scheduling applications are shown in Tables 1 and 2. These tables suggested that there are two broad categories; those that prioritize ease of use and those that prioritize detail. Particularly, we assume that recognition of symbols (○/△/×) on the coordination tool may differ from person to person. Such differences in alternative choices also may cause the possibility of difficulty in answering. Thus, we set the checking items as usability differences of answering items and interfaces and other subjective impressions. Based on this, the evaluated forms are shown in Figure 1. Each form was designed with a unique symbol-selection method, which includes radio buttons, a pull-down menu, and a different layout of comment fields. A few visually impaired computer users tested the accessibility of these forms using various screen readers such as PC-Talker, Jaws, NVDA, and Windows Narrator. In these forms, circles (○/◎) and crosses (×) represent available and unavailable for participation, respectively. In Japanese conventions, ◎ implies a higher desired order or priority than that of ○. Triangle (△) denotes partially available, with multiple meanings such as partially available for participation, adjustable but reluctant to spare time, available but can attend other meetings simultaneously, and others.

Table 1. Surveyed web-based scheduling applications.

Tool name	Feature	URL
Choseisan	Schedule adjustment tool used by 25 million users	https://chouseisan.com
Densuke	Tool featuring functions such as email notifications	https://www.densuke.biz
Tonton	Tools for scheduling mainly by mouse drag-and-drop operation	https://tonton.amaneku.com
LINE Schedule	Easy to manage with one function of the message application named LINE	–

Table 2a. The functions of web-based scheduling applications to make a schedule.

Function	Choseisan	Densuke	tonton	LINE Schedule
Setting items	• Event Name • Event Details • Candidate Date • Answer choices	• Event Name • Detailed Comment • Candidate Date • Password • Answer choices	• Event Name • Event Details • Candidate Date	• Event Name • Event Details • Candidate Date

Function	Choseisan	Densuke	tonton	LINE Schedule
Designation method of candidate date	- Specification of date and time - Text Free Descriptive	- Specification of date and time - Text Free Descriptive	Specification of date and time (Mouse operation)	Specification of date

Table 2b. The functions of web-based scheduling applications to record an appointment.

Function	Choseisan	Densuke	tonton	LINE Schedule
Answer options	○/×/△	◎/○/×/△	Hope to join / Available / Undecided / Unavailable	○/×/△
Permission to change registered attendance	Changeable even to someone’s property	Changeable even to someone’s property	Changeable only for own registration	Changeable only for own registration
Settings for registration deadlines	Unimplemented	Unimplemented	Unimplemented	Unimplemented
Prevention of duplicate registration	Unavailable	Unavailable	Unavailable	Available

Table 2c. The functions of web-based scheduling applications to aggregate appointments.

Function	Choseisan	Densuke	tonton	LINE Schedule
Tabulation result screen	Number of people and highlight the date and time with the most participants	Graphic (Timeline)	Number of people and highlighting	Tabulation result screen
CSV download	Available	Available	Unavailable	Unavailable

The features of the forms shown in Figure 1 are as follows:

- Form 1: radio button from “○/△/×” for each candidate date and a comment field
- Form 2: radio button from “○/△/×” with corresponding comment box for each candidate date

- Form 3: drop-down box with "Available to join," "Pending," "In preparation," and "Unavailable" for each candidate date and a comment field
- Form 4: radio button from "◎/○/△/×" and comment box for each candidate date

Fig. 1. Appearance of evaluated scheduling forms.

Evaluation Procedure

We asked the participants to use their personal computers to stimulate schedule registration on the scheduling web interfaces (Figure 1). Then, we asked them to evaluate the ease of registering for the schedule on the forms and adjusting with their simulated schedules. These simulated schedules include dates and corresponding plans. The participants were then asked to follow the simulated schedule to determine the date and time that they registered for availability. In addition to the content and time of the appointment, the schedule included the persona (age, occupation, etc.) of the simulated person. The scheduling situations were as follows:

1. Whether they attend multiple meetings at the same time when overlapping meetings are held online
2. When a meeting overlaps with a personal schedule, such as a part-time job, which takes precedence
3. When it is necessary to secure work time to meet deadlines, whether or not to attend or absent the time when it is not scheduled

After the registration, they were asked to answer the questionnaire described below.

Overview of the Questionnaire

The overview of the questionnaire items was as follows:

- Basic information (Gender, age, degrees of disability, personal characteristics (Big Five personality traits) obtained from the Ten Item Personality Inventory Japanese [TIPI-J] version) (Oshio et al. 52).
- Method and procedure for arranging the schedule, such as communication with others when arranging schedules, the means used for scheduling, and the participants' way of thinking about travel time.
- Usability of the scheduling interfaces, including the Usability Metric for User Experience UMUX-lite (Sauro), a simple scale consisting of two items that have been reported to correlate highly with the System Usability Scale (Lewis et al. 2102).

Analysis

After a simple tabulation of the responses to each item, we first examine the trends of the answers. Multiple regression analysis was used to examine the significant effects among the elements including the type of forms and Big Five personality traits to examine for the usability score obtained using UMUX-lite. We also checked the significance of the usability score to the forms' interface elements like the style of choices (radio buttons or a pull-down menu), the number of choices (3 or 4), and comment boxes (by date or all together). Then, we examined and discussed the significance of each element. Following that, we analyzed the contents registered by the participants and discussed the tendency to input.

Results and Discussion

Usability of Modeled Schedule-registration Forms

Figure 2 shows the results of the UMUX-lite evaluation. The results show that form 2, the form with radio buttons for "○/△/×" and a comment field for each candidate date, is likely to yield higher values. Multiple regression analysis of the differences in UMUX-lite scores for the personality and forms revealed that scores were marginally higher for form 2 with significance ($p = 0.07 < 0.10$). There is no significant difference for age, and disability conditions ($p > 0.05$). When compared by form element, individual comment boxes tended to be rated significantly higher ($p = 0.07 < 0.10$), but pull-down menus were rated significantly higher than radio buttons ($p = 0.04 < 0.05$). Some participants commented that they preferred the pull-down menus, where no circulation between items occurs. Although there was no significant difference in the number of choices ($p = 0.14 > 0.10$), only one participant preferred the combination of "◎/○/△/×," another preferred the combination of texts, and the remaining twelve persons preferred the combination of "○/△/×," according to the participants' comments.

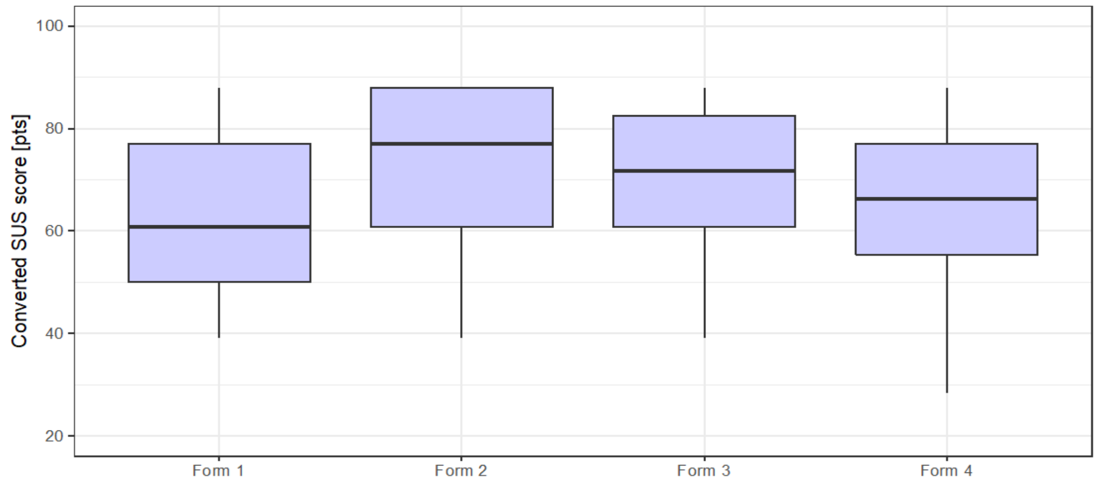


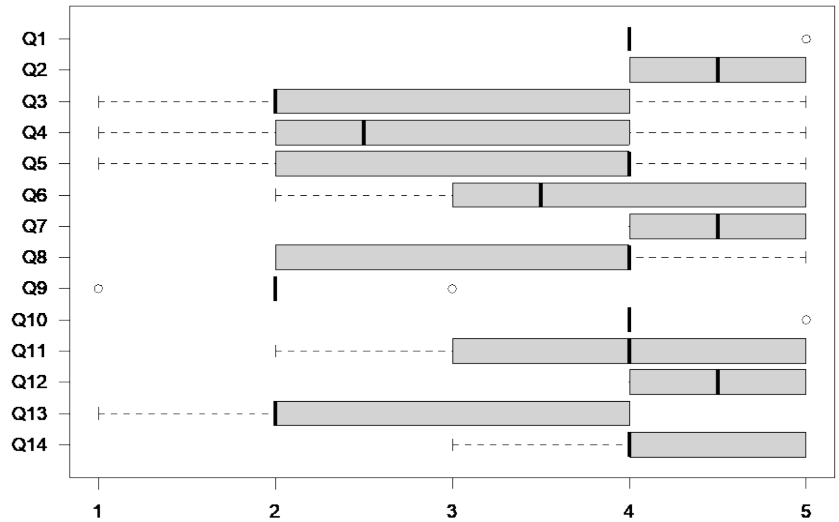
Fig. 2. Evaluated usability scores of the forms using UMUX-lite.

The fact that form 2 has a comment field for each candidate date allows the user to enter detailed information in text form, which may explain why it is rated higher than the other forms. Screen reader users also commented that this style would facilitate the identification of breaks between items. However, if the form only had one comment field, like form 1 and 3, the score will not have been higher because of the complexity of writing comments about multiple dates. Meanwhile, form 4 has a comment field for each date and time, but the need to select from four different types of symbols may have lowered the usability evaluation value. In contrast, no significant difference in usability scores was observed for the number of choices, suggesting that an interaction or an artifact of various settings may have occurred. This could be due to the ambiguity in the use of the "◎" and "○" symbols. Moreover, the possibility of higher usability scores for pull-down menus was indicated, while the evaluation values for form 3 was not significantly higher. Thus, the function to add comments for each date may have more impact on the usability score than pull-down menus. Still, it is possible that certain artifacts may have affected the results, and a detailed evaluation of each combination of interface elements is necessary.

Regarding the Big Five personality, those with higher scores for conscientiousness score significantly higher ($p < 0.05$), while those with higher scores for neuroticism score significantly lower ($p < 0.05$). Thus, there may be a personality that prefers or does not prefer the schedule adjustment form itself.

Perceived Usefulness of a Web-based Scheduling Application

Figure 3 shows the ratings for web-based scheduling tools. In this graph, the items “Q1. Easy to register my available date and time,” “Q2. Convenient to register on the web,” “Q7. Easy to reduce the frequency of communication with organizers,” “Q10. Easy to coordinate my schedule,” “Q12. Easy to reduce the frequency of contacting the coordination target,” and “Q14. Overall easier schedule compilation” were highly rated by the participants. The visually impaired were reluctant to contact via e-mail, according to their comments. However, the participants rated low in the items of “Q3. Because I can register on the web, I worry about privacy” and “Q9. Concerned about seeing other people’s schedules.” Therefore, the participants evaluated the web-based scheduling app based on the simplicity to manage the schedule and decreased communication frequency. Participants also have a few privacy concerns when it came to sharing dates and times. This tendency is the same in the visually impaired and the sighted.



Questions

Q1. Easy to register my available date and time
Q2. Convenient to register on the web
Q3. Because I can register on the web, I worry about privacy
Q4. Good to be able to concentrate on registering my available date and time
Q5. Difficult to tell the order of preference for each date and time
Q6. Easy to fill in the form with multiple candidate dates
Q7. Easy to reduce the frequency of communication with organizers
Q8. Good to register a schedule while checking others' schedules
Q9. Concerned about seeing other people's schedules
Q10. Easy to coordinate my schedule
Q11. Easy to check the availability of others simultaneously
Q12. Easy to reduce the frequency of contacting the coordination target
Q13. Increased inquiries to the coordinator regarding the operation of the coordination tool
Q14. Overall easier schedule compilation

Figure 3: Ratings for web-based scheduling tools (5: Strongly Agree ⇔ 1: Strongly disagree)

Content that the Participants Registered

Differences were observed in the symbols and comments registered depending on the date and time of the appointment to be adjusted, as well as the type of overlapping events. For example, when a private appointment overlapped with the date and time of the online meeting, the participants answered "△" and entered the following response: "I can attend if it is OK that I am late" and "I will check if I can move my private schedule."

Meanwhile, most of the participants entered "×" if the date and time of the scheduled meeting in the laboratory (online) and the date and time of other online meetings could overlap. However, some participants left comments such as "I can attend for an hour" to indicate that they could attend for a part of the time. A similar tendency was observed when a part-time job schedule overlapped with the scheduled adjustment date and time. From the above, there are differences in the responses according to the compulsory nature of the events.

Summary and Future Work

To clarify the accessibility issues of the web-based scheduling interfaces for visually impaired persons, we created four types of modeled-scheduling interfaces and then asked screen reader users to register a simulated schedule. Our achievements are as follows:

- A1. The scheduling form should possess the radio buttons for circle (○), triangle (△), cross mark (×), and a comment field for each candidate date. However, it is generally easy for screen reader users to use the pull-down menu for registering the schedule.
- A2. Visually impaired participants evaluated the web-based scheduling app on the simplicity to manage the schedule and the decrease in communication frequency. In terms of sharing dates and times, they did not have much of an issue with privacy. Regarding usability score, the persons with higher conscientiousness and lower neuroticism in the Big Five tended to score significantly higher.

Our future work would be as follows:

- Further analysis about the preference of scheduling on the teleworking conditions.
- Evaluation of the visually impaired workers.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Numbers JP18H01047, JP21H00915, JP21H05347, and JP21H00885. We would also like to thank all those who cooperated in the experiment.

Works Cited

Carvalho, Michael Crystian, et al. “Accessibility and Usability Problems Encountered on Websites and Applications in Mobile Devices by Blind and Normal-Vision Users.” Proceedings of the 33rd Annual ACM Symposium on Applied Computing, 2018, pp: 2022-2029, https://doi.org/10.1145/3167132.3167349.

Ghidini, Eduardo, et al. “Developing Apps for Visually Impaired People: Lessons Learned from Practice.” 2016 49th Hawaii International Conference on System Sciences (HICSS), pp: 5691-5700, 2016, https://doi.org/10.1109/hicss.2016.704.

Lewis, James R., et al. “UMUX-LITE: When There's No Time for the SUS.” Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2013, pp. 2099-2102., https://doi.org/10.1145/2470654.2481287.

Motomura, Shinichi, et al. “Meeting Arrangement System Based on Mobile Agent Technologies.” IPSJ Journal, 46(12):3123-3126 (2005). (in Japanese)

Oshio, Atsushi, Shingo Abe, Pino Cutrone. “Development, Reliability, and Validity of the Japanese Version of Ten Item Personality Inventory (TIPI-J).” Japanese Journal of Personality, 21(1):40-52, (2012)

Sauro, Jeff. “Measuring Usability: From the SUS to the UMUX-Lite.” MeasuringU, 2017, https://measuringu.com/umux-lite/. (Accessed: November 30, 2022)

THE JOURNAL ON TECHNOLOGY AND PERSONS WITH DISABILITIES

You Described, We Archived: A Rich Audio Description Dataset

Charity Pitcher-Cooper[1], Manali Seth[2], Benjamin Kao[2], James M. Coughlan[1], Ilmi Yoon[2]
Smith-Kettlewell Eye Research Institute[1]
Computer Science, San Francisco State University[2]
cpc@ski.org, mseth@mail.sfsu.edu, bkao1@mail.sfsu.edu, coughlan@ski.org, ilmi@sfsu.edu

Abstract

The You Described, We Archived dataset (YuWA) is a collaboration between San Francisco State University and The Smith-Kettlewell Eye Research Institute. It includes audio description (AD) data collected worldwide 2013-2022 through YouDescribe, an accessibility tool for adding audio descriptions to YouTube videos. YouDescribe, a web-based audio description tool along with an iOS viewing app, has a community of 12,000+ average annual visitors, with approximately 3,000 volunteer describers, and has created over 5,500 audio described YouTube videos. Blind and visually impaired (BVI) viewers request videos, which then are saved to a wish list and volunteer audio describers select a video, write a script, record audio clips, and edit clip placement to create an audio description. The AD tracks are stored separately, posted for public view at https://youdescribe.org/ and played together with the YouTube video. The YuWA audio description data paired with the describer and viewer metadata, and collection timeline has a large number of research applications including artificial intelligence, machine learning, sociolinguistics, audio description, video understanding, video retrieval and video-language grounding tasks.

Keywords

Video Accessibility, Blind and Low Vision, Audio Description, Artificial Intelligence, Machine Learning, Sociolinguistics.

Introduction

Thousands of videos are uploaded to online video platforms daily; unfortunately, the visual content of these videos is inaccessible to blind and visually impaired (BVI) individuals and a lack of accessible video content has a profound negative impact on learning and community connections (Packer et al.). According to the World Health Organization, at least 2.2 billion people globally are visually impaired (World Health Organization). Over eight million Americans (3.3 % by population) are visually impaired and may rely on a screen reader, screen magnifier or have a form of color blindness (“Accessibility Statistics | Interactive Accessibility,” “53 Web Accessibility Statistics [Updated for 2022]”).

The video platform YouTube is the second largest search engine with 500 hours of video content uploaded every minute, and 5 billion videos watched per day (“30 Eye-Opening YouTube Facts”). It is a major video content source for: Film & Animation, Music, Autos & Vehicles, Travel & Events, Pets & Animals, Sports, People & Blogs, and Gaming. An effective way to bridge the video accessibility gap is to add audio descriptions, an additional narration track providing visual information unable to be inferred from audio cues alone such as setting, facial expressions, gestures, on-screen text, style of dress, or any other relevant information, synchronized with the video that can be turned on or off as needed.

YouDescribe (YD) is a unique, free, crowdsourcing tool for adding audio descriptions to YouTube videos (“YouDescribe - Audio Description for YouTube Videos”). It was first launched in 2013 as a project of The Smith-Kettlewell Video Description Research and Development Center by scientist Dr. Joshua Miele. The current version of YouDescribe has been in continuous use since a major renovation in 2017. With the accelerating inclusion of short-form (rapidly produced, under 5 minute) videos via social media, the need for timely AD has only become more critical. Audio description made with YouDescribe is a complementary service to professional quality AD; it strives to be accurate, practicable, timely and fun. While most recent movies and TV shows have professional AD available, AD for home movies, educational videos for work and school, and short-form content is in very high demand.

Discussion

Audio description is an art that requires critical thinking about content importance, as well as a precise vocabulary to create harmonious descriptions with the fewest number of syllables to fit the space imposed by comprehension-critical soundscape and dialog. More than crowdsourcing, YouDescribe is a community of people who want accessible video, combined with volunteers who have a vested interest in creating useful content. At YouDescribe, viewers have an active role in requesting content that is important and enjoyable to them. The YouDescribe AD tool focuses on functionality for individual viewers and ease of use for volunteer describers. Training is encouraged for all new audio describers through in-person or virtual training provided by: The Smith-Kettlewell Eye Research Institute, a formal class provided by a describer organization, college class, apprenticeship, or through YouDescribe's accessible, text-based tutorials with a complimentary educational YouTube channel ("Accessibility: A Guide to Building Future User Interfaces"; "Literature and Disability"; "The Art of Writing"; "YouDescribe - Audio Description for YouTube Videos"; "YouDescribe - YouTube"). While it is possible to use YouDescribe through trial and error, the text-based training materials are brief and cover both how to use YouDescribe as well as standard industry-wide AD guidelines. The tutorial page is one of the most common visitor-accessed pages, second to the landing page. Visitors are provided step-by-step instructions on where the control buttons are located, how to request wish list items, essential describer training and general troubleshooting. The Smith-Kettlewell Eye Research Institute has sponsored training since 2013 with inventor Joshua Miele, Professor Yue-Ting Siu and trained audio describer Charity Pitcher-Cooper. Populations brand new to the concept of audio description are generally able to complete their first 3-minute AD, on a video with content well known to them in about 1.5 hours and describers already familiar with AD take about an hour for their first description made with YouDescribe (Pitcher-Cooper and Brabyn). Experienced describers often choose research-heavy projects like AD for museum pieces or full TV shows and feature films requiring many hours of background study, as well as lengthy recording time.

The You Described, We Archived (YuWA) dataset protects the privacy of the viewer and describer community and provides only anonymized data. For the purposes of this paper, we are concentrating on metadata for visitors and volunteers who used the current YouDescribe web interface officially launched May 17th, 2017. Audio descriptions made before Spring, 2017 were created with a similar but not identical interface launched in 2013. That interface became unstable for use and was completely rebuilt in 2017. The YuWA data repository includes all YouDescribe related AD from 2013-2022 and can be sorted to include or exclude important YouDescribe milestones.

Google Analytics for https://youdescribe.org was implemented in July, 2020 and reflects visitor trends for the past two years (Fig. 1). Of the 35 thousand visitors, 65% of visitors are from the US with notable traffic from Canada, United Kingdom, and Australia (Fig. 2). Around 14,000+ users have accessed the website directly, 11,000+ were referral traffic (the segment of traffic that arrives at YouDescribe from a link on another domain), and 11,000+ users visited via Organic Search, Social or Email combined (Fig. 3). Describers range in age from 13 and above. Approximately 40% of the web app traffic is male, and 60% is female (Fig. 4). (These statistics are a general overview of visitor trends; Google Analytics for age and gender isn't available for all visitors and not all collected data is accurate; for example, those under 18 have no tracked data, and collected account information may not align with a visitor's gender). Our largest number of audio descriptions, 5,895 videos out of a total 6,484, are recorded in English; language tags for English include Australia, Belize, Canada, Ireland, Jamaica, New Zealand, South Africa, Trinidad, United Kingdom, and the United States. The most compatible browser for the YouDescribe website is Google Chrome. As of August 2022, Chrome is the leading internet browser in the world with a global market share of 65.52% ("Most Popular Web Browsers in 2022 | Oberlo"). YouDescribe visitors also enjoyed viewing in Safari (Apple), Edge (Microsoft), and Firefox (Mozilla).

Fig. 1. Total, New and Returning Visitors (07/2020-09/2022, Google Analytics).

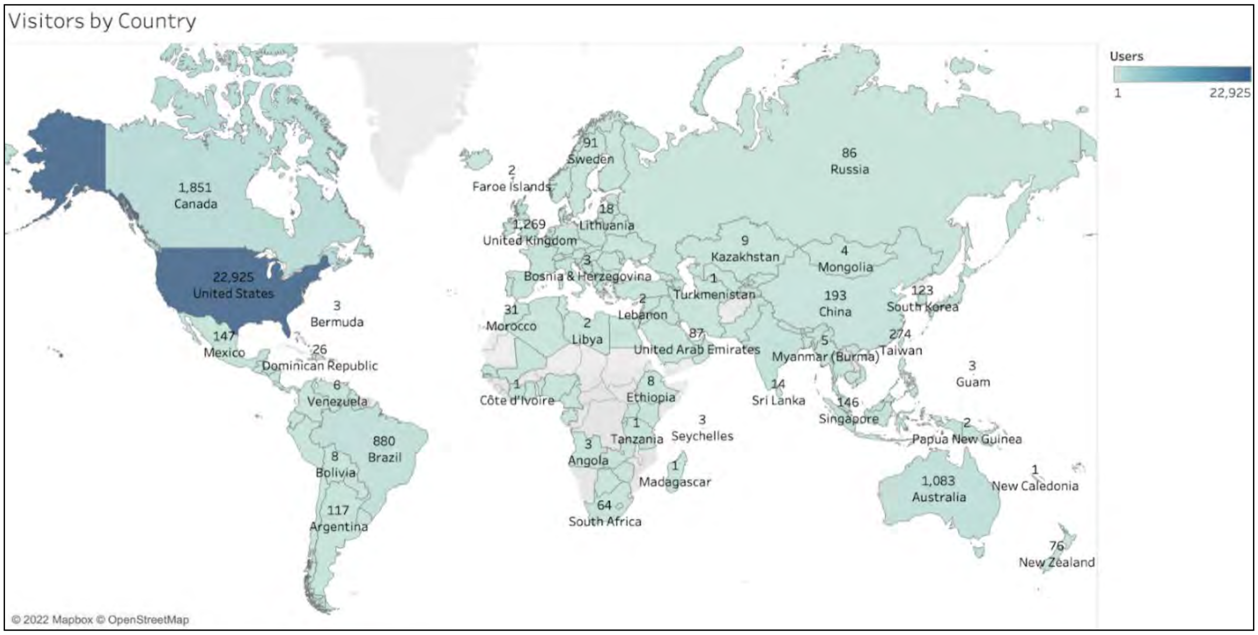


Fig. 2. Choropleth map of visitors by country (07/2020-09/2022, Google Analytics).

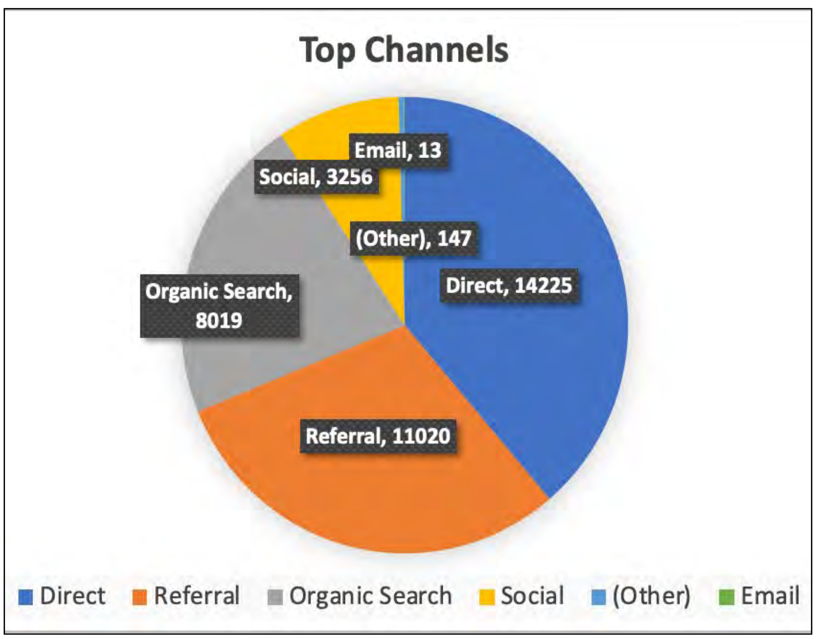


Fig. 3. Top Channel Traffic for YouDescribe (07/2020-09/2022, Google Analytics).

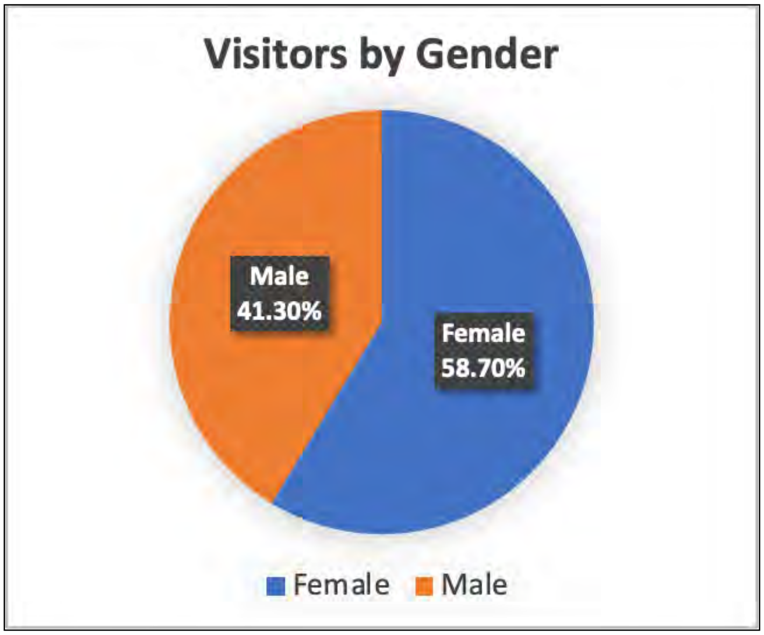


Fig. 4. Visitors by Gender (07/2020-09/2022, Google Analytics).

Our wish list, a curated list of what videos viewers want to have described, weighted for when the video was requested and how many “up-votes” each video has (videos requested multiple times go to the top of the queue), is one of the most important features to understand the needs of the community. All YouDescribe trainings encourage volunteers to choose videos from the wish list. The maximum number of times that a video is requested is 32. The most requested video category is ‘People & Blogs’ with about 500 requests, followed by 450+ for ‘Entertainment’ and 400+ for ‘Film & Animation’. In addition to adding videos to the wish list, viewers rate the quality of the AD on a one-to-five-star basis (1 being poor, 5 being excellent) and those ratings are matched to the describer’s Google ID and posted publicly alongside of the description. In the case of multiple descriptions for the same YouTube video ID, the highest rated AD is listed first. In addition to the public facing ratings, viewers can select feedback from a list to help new describers with specific improvements on: audio quality, diction, balance of inline and extended tracks, a need for less or more description, voice tone matching the tone of video, description given before action/audio cues, and a lack of description for onscreen text. About 21% of the AD has been rated by viewers, and 79% of videos remain unrated (Fig. 5). Of the videos rated, 59% are excellent, 25% are very good, 8% are good, 5% are fair, and only 3% are rated as poor. The quality and utility of the YuWA dataset would be improved with a greater number of videos being rated by the BVI community.

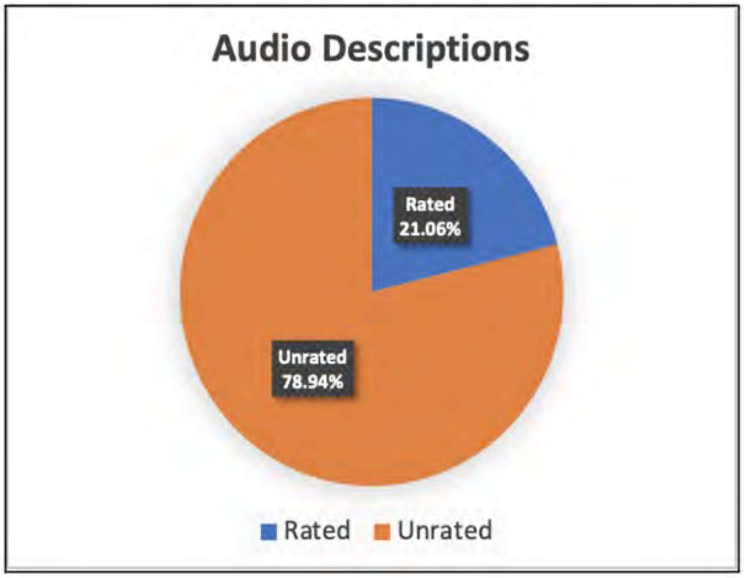


Fig. 5. Audio Descriptions rated and unrated (03/2017-09/2022, Live Dataset).

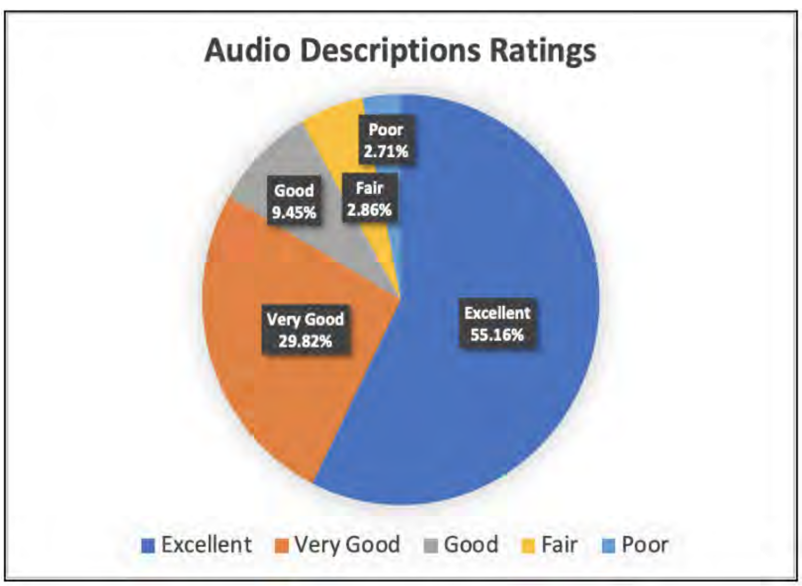


Fig. 6. Audio Descriptions ratings: excellent to poor (03/2017-09/2022, Live Dataset).

As of Sept. 2022, there are 3,000+ describers, 6,400+ audio descriptions with approximately 1,000+ ADs added in 2022, and on average 90+ videos described each month, depicting a general trend of more volunteers providing descriptions each year. The dataset covers a vast domain of videos in 32 titled categories including the 15 most popular: Film & Animation, Music, Autos & Vehicles, Travel & Events, Pets & Animals, Sports, People & Blogs, Gaming, Comedy, Entertainment, How-To & Style, News & Politics, Nonprofits & Activism, Education, and Science & Technology. YuWA has 5,500+ videos with a total duration of about 310 hours and an average video duration of about 5.5 minutes. Currently there are 76,000+ audio clips (about 10 audio clips per 5-minute description with about one audio clip per 30 seconds). The audio clips are transcribed with Listen by Code and Google Cloud's Speech-to-Text API. As speech to text services improve, the dataset may be updated to reflect those changes. Similarly, as new YouDescribe AD content is published, additional data may be added to the YuWA database.

The following are some specialty datasets that are of interest to researchers in various fields, highlighted to showcase the possible utility of audio description across disciplines. YuWA includes a specialized dataset for 75+ descriptions of an identical video made for Carnegie Mellon's Human Computer Interaction Institute course. "Students in Special Topics: Accessibility: A Guide to Building Future User Interfaces" have uploaded the video to their own YouTube Channel (each video has the same content but a different YouTube ID), and described an employment graph. The shortness of the video (21 seconds), along with the 75 aggregate descriptions collected over 5 years has a number of possible study applications in audio description research, comparative linguistic studies, or to test against computer generated descriptions with multiple trainers. While other videos at YouDescribe have multiple descriptions, they are listed in order of quality rating, and attached to a single YouTube ID making them easier to be cataloged.

We also have a premium set of describer data that can be compared to the full dataset, or future describer datasets. The audio description in this specialized set was done exclusively by describers aged 13-18, earning community volunteer hours for a Northern California service organization. A 1–2 hour training was supplied by a Smith-Kettlewell sponsored trainer, and club members received personalized and general feedback over the year from trainer Charity Pitcher-Cooper as desired. Describers had additional support from an adult coordinator who was also present for the describer trainings. More experienced student describers (1+ years) supported new describers with their audio descriptions. Many describers from this group continued creating AD after age 18, and their post service club member descriptions are included. Because YouDescribe is used all over the world, in any language, and our volunteers have a variety of different describer backgrounds, this specialized data subset offers a known age range, and audio description training cohort that can be compared to other describers.

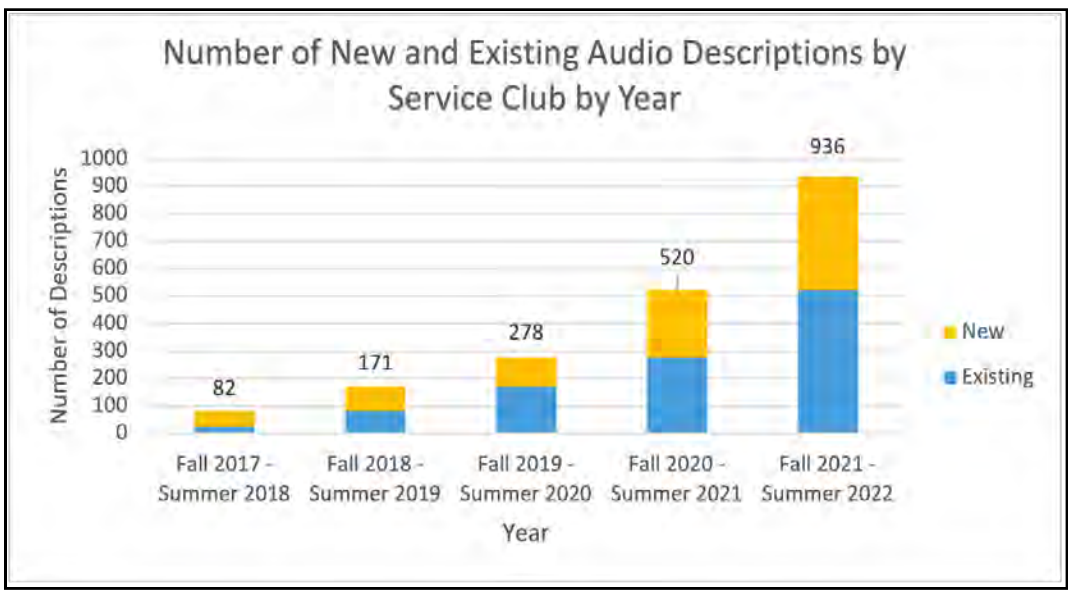


Fig. 7. Audio Descriptions by Service Club (Fall 2017 - Summer 2022, Premium Dataset).

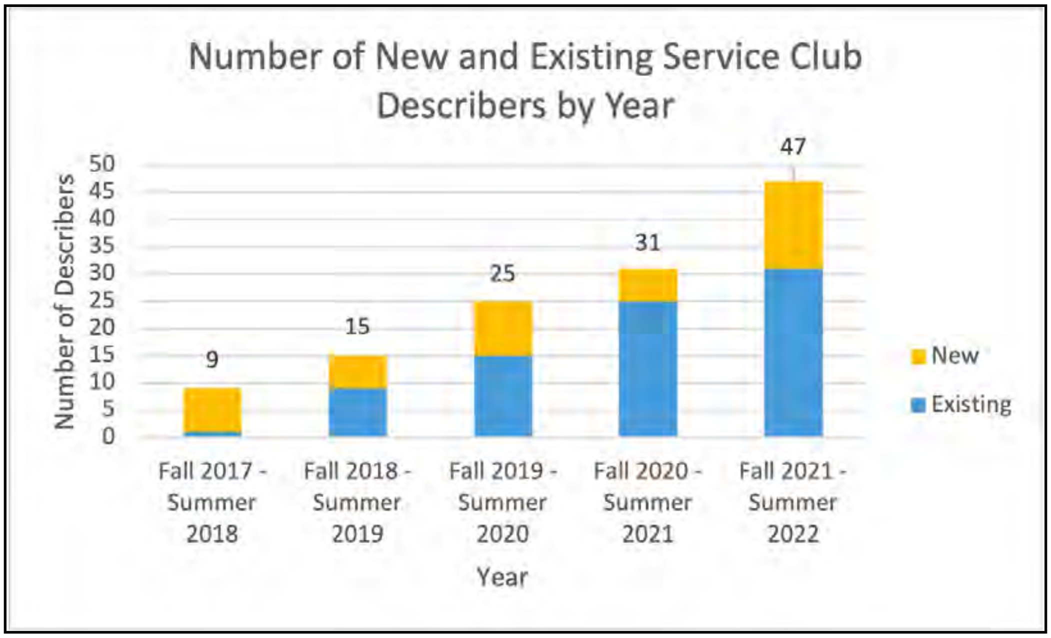


Fig. 8. Service Club Describers (Fall 2017 - Summer 2022, Premium Dataset).

Conclusions

High-quality datasets that can effectively contribute to advancing artificial intelligence (AI) and deep learning (DL) on video understanding, video retrieval, or video-language grounding tasks are increasingly valuable as computation power and algorithms grow in power. The YuWA dataset has many possible applications given the large volume of data collected over the past 9 years. For example, The Movie Audio Description (MAD) scalable dataset which focuses on video-language grounding tasks was presented by King Abdullah University of Science and Technology (KAUST) (Soldan et al.). The MAD database and the YuWA dataset share a similar description origin with respect to the crowdsourcing of data. Crowdsourced data is considered "noisy," as humans make mistakes and recording instruments can sometimes be inaccurate, the data collected has some error bound to it. Noisy data can significantly impact the prediction of any meaningful information. MAD attempted to overcome the hidden biases often found in video-language grounding tasks and increase the generalization capabilities from visual features for long-form video, aiming to have high-quality temporal localization of detecting activities to output beginning and ending timestamps of the language in untrimmed videos. Ultimately, they compared a model trained on the much smaller, quality-controlled LSMDC-G training set (which is only 32% the size of the MAD training set), with the same model trained on the full MAD dataset that yielded a relative quality improvement of 20%. As a control, when the MAD model was trained on just a small, 32% sub-section of the MAD training set, it performed worse than the LSMDC-G training set. In this case, MAD researchers showed the success of using a large, diverse, crowdsourced dataset, like YuWA, and achieved better results than smaller, less noisy, datasets.

The Visual Geometry Group from the University of Oxford introduced QuerYD, an audio description video dataset, for performing retrieval and event localization in videos based on a subset of YouDescribe data (Oncescu et al.). Oncescu et al's training research illustrated that the content descriptions made with YouDescribe for the BVI community are more relevant than dialogue alone and more detailed than previous description attempts, which can be observed to contain many superficial or uninformative descriptions. The QuerYD dataset has been influential for a number of applications and is cited in ten scholarly publications. However, QuerYD focuses on the AD speech to text outputs without the inclusion of metadata on viewers or describers which greatly limits its scope for applications specifically for audio description research, and sociolinguistic applications. In contrast, the YuWA open-source dataset has two audio tracks which include the original audio and high-quality audio descriptions provided by the volunteers without any background disturbance with a precise and direct relationship to video segments. It provides audio narrations which are much more detailed and relevant than the standard narration transcriptions used by other ML training datasets. This is possibly due to the addition of an extended audio clip tool. Due to the application to scholarly and work critical audio descriptions, the YD tool has two kinds of track styles: inline (which is played over the video soundtrack in dialog and soundscape critical pauses) or extended (where the source video is paused, allowing for a longer description to be inserted). Professional AD does not utilize extended track timing in order to keep pace with the video content. Because YouDescribe's purpose includes detail-oriented fan-base observations, as well as incredibly content-critical educational, and work-related AD, describers have the option to record longer, more detailed tracks with the extended play mode. Keeping in line with international audio description standards, describers are urged to use the greatly preferred inline track style as much as possible and extended only when necessary. The large volume of data, plus the dedication of the volunteers, combined with the structure of the tool itself (the addition of an extended audio track) makes the YuWA dataset both unique and versatile. This organized, curated, open-source dataset will most certainly be used for AI-based audio description projects.

Currently, a project headed by San Francisco State University computer science program called YouDescribeX seeks to generate audio descriptions either automatically or semi-automatically by making use of the Human-in-loop Machine Learning (HILML) approach to video description, automating video text generation and scene segmentation and then allowing humans to edit the AI generated output (Bodi, et al). A comparison of YouDescribeX AD to the YuWA dataset has the near-future potential to build a user-friendly AD tool, and provide audio descriptions for a much larger number of YouTube videos than would be possible with human-generated audio descriptions alone.

Moreover, the YuWA dataset has included detailed, anonymized metadata as well as useful summary statistics allowing for potential sociolinguistic study of topics such as describer register (a variety of language used for a particular purpose or in a particular communicative situation) and viewer satisfaction; such metadata could enable the creation of AD-specific word selection guidelines helping both humans and computers find the most descriptive phrases with the shortest number of syllables, combined with impeccable track placement.

The YuWA Dataset described here is hosted in the You Described, We Archived repository at https://github.com/youdescribe-sfsu/You-Described-We-Archived

Acknowledgments

This work was supported by NEI- R01EY020925, 2010-2014; Department of Education-H327J110005 2011-2013; NIDLIRR- RERC on Blindness and Low Vision- 90RE5024 2016-2022 and 90REGE0018 2022-2027; Ability Central (formerly The Disabilities Communication Fund)- Improving Access to Described Video 2016-2017, Training the Next Generation of Audio Describers 2017-2018, and Machine Learning Enhanced Video Accessibility for the Blind 2020-2022.

Works Cited

"30 Eye-Opening YouTube Facts, Figures and Statistics You Should Know in 2022." *Cloud Income* 17 Apr. 2020, https://cloudincome.com/youtube-statistics/#youtubestats.

"53 Web Accessibility Statistics [Updated for 2022]." *Don't Do It Yourself*, 24 Mar. 2022, https://ddiy.co/web-accessibility-statistics/.

"Accessibility Statistics | Interactive Accessibility." *Interactive Accessibility*, www.interactiveaccessibility.com/accessibility-statistics.

Alayrac, Jean-Baptiste, et al. "Flamingo: a visual language model for few-shot learning." arXiv preprint arXiv:2204.14198 (2022).

"Audio Description Basics for Beginners: A Do's and Don'ts Video Tutorial. - *YouTube*." www.youtube.com, www.youtube.com/playlist?list=PLNJrbI_nyy9uzywoJfyDRoeKA1SaIEFJ7.

Bodi, Aditya, et al. "Automated Video Description for Blind and Low Vision Users." *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 8 May 2021, 10.1145/3411763.3451810.

Carnegie Mellon University, Human Computer Interaction Institute. "Accessibility: A Guide to Building Future User Interfaces" HCI Undergraduate: 05-499, HCI Graduate: 05-899, 2017-2022. https://www.hcii.cmu.edu/course/special-topics-accessibility-guide-building-future-user-interfaces

"Lecciónes de Descripción - *YouTube*." www.youtube.com, www.youtube.com/playlist?list=PLNJrbI_nyy9sZmflPvSFv_WwuNVivLyBr. Accessed 6 Oct. 2022.

"Most Popular Web Browsers in 2022 [Sep '22 Update] | *Oberlo*." www.oberlo.com, www.oberlo.com/statistics/browser-market-share#:~:text=As%20of%20August%202022%2C%20Google.

Oncescu, Andreea-Maria, et al. "QUERYD: A Video Dataset with High-Quality Text and Audio Narrations." ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 6 June 2021, 10.1109/icassp39728.2021.9414640.

Packer, Jaclyn, et al. "An Overview of Video Description: History, Benefits, and Guidelines." *Journal of Visual Impairment & Blindness*, vol. 109, no. 2, Mar. 2015, pp. 83–93, 10.1177/0145482x1510900204.

Pitcher-Cooper, Charity, and John Brabyn. "The Unstoppable Utility of YouDescribe." Wireless RERC 2021 State of Technology Forum Proceedings, 30 Sept. 2021, https://www.ski.org/unstoppable-utility-youdescribe-wireless-rerc-2021-state-technology-forum-proceedings.

Soldan, Mattia, et al. "MAD: A Scalable Dataset for Language Grounding in Videos from Movie Audio Descriptions." 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2022, 10.1109/cvpr52688.2022.00497.

University of California, Berkeley. *Online Course Catalog*, English 165. "The Art of Writing: The Visible Made Verbal" Spring 2019 https://english.berkeley.edu/courses/5931.

University of California, Berkeley. *Online Course Catalog*, English 175. "Literature and Disability." FALL, 2020, Fall 2021 https://english.berkeley.edu/courses/7472.

"What Is YouDescribe? A Video Tutorial - *YouTube*." www.youtube.com, www.youtube.com/playlist?list=PLNJrbI_nyy9sjqZ-Wcn6sX868i9KtdNrT.

World Health Organization. "Blindness and Vision Impairment." *WHO International, World Health Organization*: WHO, 11 Oct. 2021, www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment.

Xu, Hu, et al. "VideoCLIP: Contrastive Pre-Training for Zero-Shot Video-Text Understanding." Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, 2021, 10.18653/v1/2021.emnlp-main.544.

Xu, Hu, et al. "VLM: Task-Agnostic Video-Language Model Pre-Training for Video Understanding." Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, 2021, 10.18653/v1/2021.findings-acl.370.

"YouCook2: Large-Scale Cooking Video Dataset for Procedure Understanding and Description Generation." *YouCook2*, https://youcook2.eecs.umich.edu/.

"YouDescribe - Audio Description for YouTube Videos." *YouDescribe*, https://youdescribe.org/.

"YouDescribe - *YouTube*." www.youtube.com, www.youtube.com/channel/UCnXTm_yhrDJhUcH9hkwdisw

Zhu, Linchao, and Yi Yang. "ActBERT: Learning Global-Local Video-Text Representations." 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 2020, 10.1109/cvpr42600.2020.00877.

Evaluation of Anonymized Sign Language Videos Filtered Using MediaPipe

Andrew Luna, James Waller, Raja Kushalnagar, Christian Vogler
Gallaudet University
andrewluna368@gmail.com, james.waller@gallaudet.edu,
raja.kushalnagar@gallaudet.edu, christian.vogler@gallaudet.edu

Abstract

This study investigates the feasibility of using MediaPipe to anonymize sign language videos. Recent research has developed techniques for anonymizing the identity of a signer in a video, while preserving the signed message. Many of these prototypes are computationally intensive and are not currently useable for everyday automated real-time use. This gap MediaPipe, a tool developed by Google for tracking body movement in video, could be feasible for real-time anonymization, but has not yet been evaluated for its feasibility in sign anonymization. We fill this gap with a study in which deaf signers (n=10) view two filters developed using MediaPipe: a face mesh filter that covers only the face with an avatar-like face mask and a silhouette filter that covers the whole body in a solid monochrome, with interconnected dots showing the skeleton of the signer. Results show that signers are adept at understanding and reproducing short sentences covered by either filter. However, the filters are described as unnatural, and signers note facial movements are limited. We conclude that MediaPipe is likely robust enough for understanding manual information in signs but not necessarily for capturing facial information, and we suggest further improvements to the two filters.

Keywords

Sign language; privacy; MediaPipe; real-time communication.

Introduction

A peculiar feature of written language, particularly typed messages, is that the message can be presented separately from the messenger: people can communicate through text without having to disclose their identity. Users leave reviews, participate in online forums, and discuss sensitive issues while remaining anonymous.

However, the messenger may want to avoid using the written modality, such as speech or sign. For example, many deaf and hard of hearing (DHH) people prefer to use sign language over spoken or written language. In the U.S., official announcements meant for deaf audiences are often in both American Sign Language (ASL) and written English. Many online DHH discussion groups and DHH social media users prefer to post ASL videos rather than written text. It is important that DHH people have an option to communicate using sign language if they want to anonymously discuss sensitive or confidential topics online.

Recent research has investigated various filters to make the signer in the video more anonymous – by covering, distorting, or substituting parts of the signing video to enhance anonymity while also (hopefully) preserving information about the message content (Bragg et al., 2020; Lee et al., 2021; Saunders et al., 2021). This task of separating information pertaining to the messenger (speaker identity) from information pertaining to the message's content is not trivial. In parallel work on voice conversion in spoken language, Sisman et al. (2021) note that many aspects of a spoken utterance, such as intonation, prosody, word choice, and sentence structure all can contribute to the listener's understanding of the message as well as identification of the speaker – making 'perfect' anonymization difficult. However, some of these factors are relatively more important for understanding the message. Sisman et al. suggest that anonymization methods look to eliminate speaker identifying information related to prosody and intonation but keep word choice or sentence structure.

For sign language anonymization, researchers must also choose what information to keep and what to remove. As a baseline condition, several teams have used a relatively straightforward 'tiger-mask' developed by Jeeliz (see Lee et al., 2021; Bragg et al., 2020 for application to sign anonymization). This approach covers the face with a static tiger face while keeping the hands visible. The mask does not keep track of face movements, except to emit bubbles when the mouth is opened. This approach operates under the assumption that the face is a crucial part of identifying people, while most of the message content itself is produced on the hands. However, as the authors acknowledge, this approach is probably too simplistic. While many humans (and facial recognition technology) are indeed adept at identifying people by face alone, there is plenty of other information available elsewhere in a video that give clues to the person's identity: the background, clothes, and body shape, and even the hands themselves. Conversely, in American Sign Language information is not conveyed through hands alone: eye gaze, mouth movements and facial expressions contribute to the message content in ASL and other sign languages (Pfau and Quer, 2013; Hermann and Steinbach, 2013; Wilbur, 2000).

An alternative strategy is to keep some or all information about how the signer moves in the video (e.g., movement of the hand, fingers, and face), while eliminating static information in the video such as background, hair, clothes, body shape, and/or color information of any kind. Both Lee et al. and Bragg et al. include such approaches in their study. Bragg et al. use a 'frame cel shading' approach (altered from Jeeliz) in which "the full frame is replaced with a flattened grayscale version”. Lee et al. use an approach that swaps the face of the signer with a novel face, while retaining the movements and expression of the original face. Saunders et al. use a similar method, ‘AnonySign’, which generates a model that retains the movement of the original signer’s face, body, and hands but with a completely different ‘person’ (with altered hair, clothes, body type, etc.).

The authors broadly show that these more dynamic approaches are more favorably received than the simple static mask filter. However, they are generally more computationally intensive and not currently available for widespread automated use. The goal for sign language anonymization is to be use-able in real time or near real time on a variety of devices – in particular mobile phones. For our study, we focused on MediaPipe API, developed at Google for tracking bodies, face, and hands. Compared to the above approaches, MediaPipe has low resource requirements and offers the potential of real-time tracking. Researchers have demonstrated that MediaPipe holds some promise for use in sign language recognition and handshape detection; we examined whether it is sufficiently accurate and detailed for *humans* (specifically, deaf signers) to easily understand the content of the message.

We developed two filters developed from MediaPipe meant to anonymize signers while keeping the signed message understandable, and tested them in a study with DHH participants.

Implementation of Filters

MediaPipe is an open-source AI tracking API developed by Google (Lugaresi et al., 2091; Zhang et al., 2020). The API identifies and tracks hands, face, body, and/or objects in video input. MediaPipe takes as input simple 2D images/videos and does not require specialized devices (such as 360-degree cameras or gloves). It also has low resource requirements and can run in real time on both personal computers and mobile devices. MediaPipe applies pre-trained models ('solutions') for tracking bodies to 2D video to output estimations of the 3D coordinates of various 'landmarks' on the face, hands, and/or torso – the face mesh solution, for example, estimates up to 468 landmarks on the face. These coordinates can then be used to generate avatars or be drawn directly onto a silhouette as dots.

The present study tested two new privacy filters that use MediaPipe. The first filter, the face mesh filter, utilizes landmark outputs from MediaPipe's face mesh solution as vertices to create a three-dimensional face mesh in the Unity Engine using the MediaPipe Unity Plugin. The face mesh covers the signer's head and captures the signer's facial expression. To ensure the signers original face is fully covered, a larger black box is generated under the face mask that covers any details surrounding the face. See Figure 1 for an example. The rest of the signer's body and video background is left unobstructed.

Fig. 1. Example of the Face-Mesh Filter.

The second filter, a silhouette filter, uses MediaPipe's holistic solution to create its output image. The holistic solution generates several sets of landmarks corresponding to the face mesh, the person's 'pose' (e.g., torso and legs), and both hands. Moreover, the holistic solution also outputs a segmentation mask that identifies the certainty each pixel is part of the 'human' or the 'background'. Following MediaPipe's own demo, we use the segmentation mask to cover the signer's body with a single color. The landmarks are then used to draw interconnected vertices representing the approximate 'skeleton' of the person. The silhouette filter uses the face, left hand, right hand, and pose landmark sets to draw the final image. Dots are created to represent the landmark locations, and lines are created to connect the dots.

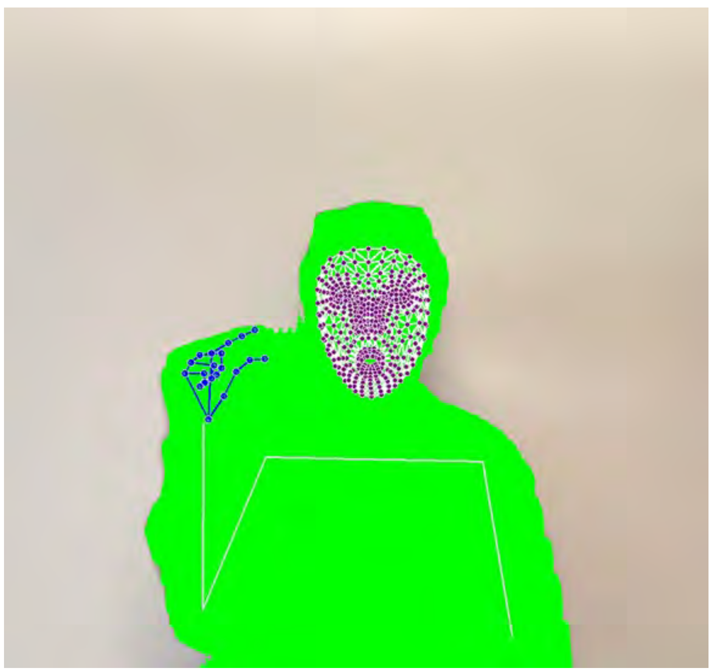


Fig. 2. Example of the Full-Body Silhouette Filter.

Previous approaches mostly focus on semi-realistic filters, like deepfakes and avatars. However, during development we observed the simple dots-and-lines approach was surprisingly understandable. Moreover, previous work in the psychology of perception show that deaf signers are able to extract significant information from relatively sparse moving dot displays (Poinzer et al., 1981; Leannah et al., 2022; Quandt et al., 2021), and so we included the silhouette filter in our study as a second condition. We predicted that compared to the face mesh filter, the silhouette filter would increase anonymity at the expense of comprehension and naturalness.

Methodology

10 deaf and hard-of-hearing adults who used ASL in their daily life took part in this study. Participants were recruited through fliers posted on campus at a university with a large DHH population and through word of mouth. Participants ranged in age from 27 to 44 years old ($M = 33.8$, $SD = 7.7$). Nine identified as Deaf and used ASL everyday while one identified as hard-of-hearing and used ASL occasionally. Seven participants learned ASL before the age of five, while two learned ASL in adolescence (12-18 years) and one learned ASL in adulthood.

At the beginning of the study, the participants were given a consent form and video release form in English. The experimenter also offered to explain any details of the consent process in ASL, and verbally let the participant know when they started the recording on Zoom. The participants then began the study, which consisted of demographic questions (hearing status, age of ASL acquisition, frequency of ASL use) followed by the main study.

For the main study we filmed short sentences produced by fluent signers and applied the filters to the sentences to create ‘anonymized’ ASL clips. For each filter, signers watched a video clip with the filter applied, as well as two baseline unfiltered clips. After watching each video, signers reproduced the sentence signed in the video from memory, to test whether they understood the sentence. After the clips were shown, participants were asked about their opinion of the two filters. They rated how understandable each filter was, how anonymous each filter was, and how easily they could infer the facial expression of the signer – for each dimension, they were given a statement and responded whether they agreed or disagreed with the statement (5-point Agree/Disagree scale: Strongly agree, Agree, Neutral, Disagree, Strongly disagree). Participants rated the following statements for both filters.

- *Anonymity:* '[This filter] is good at hiding the signer's identity.'
- *Understandability:* 'The video of [this filter] was understandable.'
- *Clarity of facial expression:* 'It is easy to see the signer's facial expression with [this filter]'

They also were asked to directly compare the two filters along two of these dimensions. They were shown the following statements and rated their agreement with the statements (5-point Agree/Disagree scale: Strongly agree, Agree, Neutral, Disagree, Strongly disagree):

- *Anonymity:* 'The green silhouette filter is more anonymous than the face mask filter'
- *Understandability:* 'The green silhouette filter is more understandable than the face mask filter'

They were also asked to pick which filter they preferred, and they had an opportunity to provide open-ended feedback on both filters.

Discussion

Comprehension of the Filters

Overall, most signers successfully reproduced the sentences they viewed even with the filter applied. Seven of ten participants reproduced the meaning of the sentences with face mesh filter applied, and eight reproduced the sentences with the green silhouette filter applied.

Filter Ratings

Participants rated each filter along three dimensions – whether the facial expressions were clear, whether the video was anonymous, and if the signing was understandable – on a scale of 1 to 5 (Strongly Disagree to Strongly Agree), as shown in Figure 3. Participants gave low ratings for clarity of facial expressions for both face-mesh filter (*M* = *2.3*), and the silhouette filter (*M* = *2.6*). However, they received higher ratings for understandability (Face-mesh – *M* = *3.8*; Silhouette – *M* = *4.1*), and for anonymity (Face-mesh – *M* = *3.7*; Silhouette – *M* = *3.9*). There were no significant differences in rating scores between the two filters for any of the dimensions ($p > 0.05$ for all pairwise comparisons).

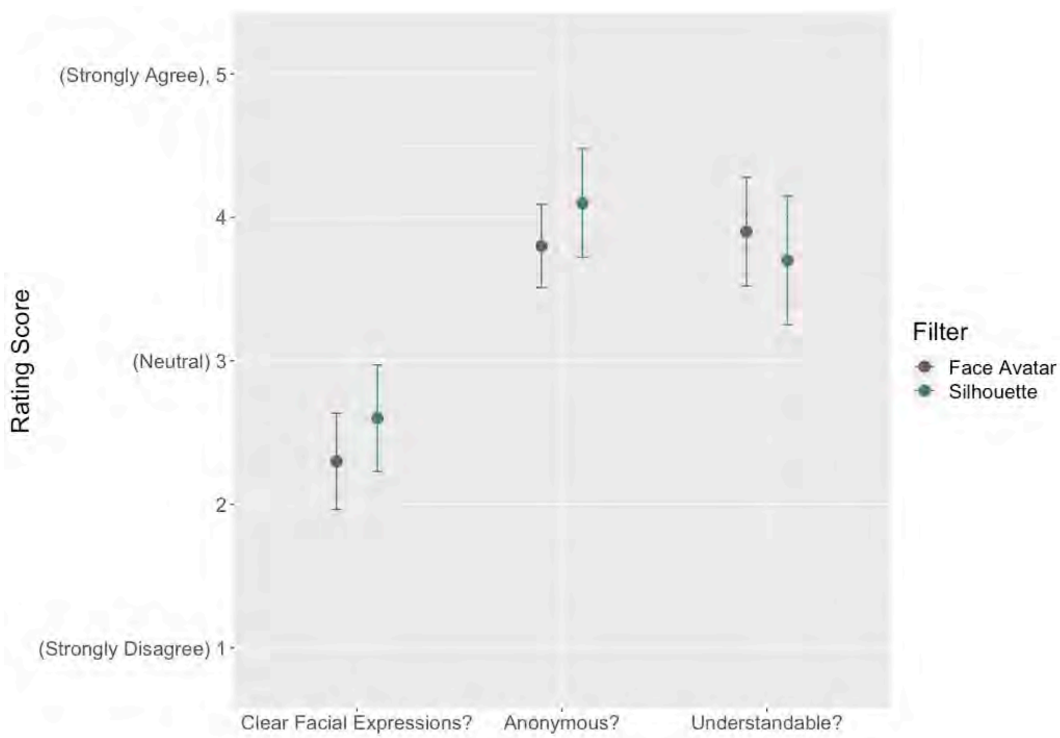


Fig. 3. Participant average ratings of clarity of facial expressions, anonymity, and ease of understanding for each of the two filters.

When asked to pick which of the two filters they preferred, 70% - seven of ten participants - preferred the silhouette filter, while 30% picked the face-mesh filter. We also asked them to compare the two filters along two dimensions – anonymity and understandability. As shown in Figure 4, participants were generally in agreement that the silhouette filter was more anonymous: only one participant disagreed with that statement, and most participants agreed or strongly agreed that the full-body filter was more anonymous – when converted to a 1 to 5 scale, the mean rating was 4.0, significantly higher than a neutral rating of 3 (one-sample t test: $t = 3.0$, $p = 0.015$). When asked if the silhouette filter was more *understandable* than the face-mesh filter, participants were split: half of participants disagreed (or strongly disagreed) and half agreed (or strongly agreed) – as shown in Figure 5. The mean rating was a 2.9, and no significant preference was shown for either filter along this dimension (one-sample t-test: $t = 0.2$, $p > 0.10$)

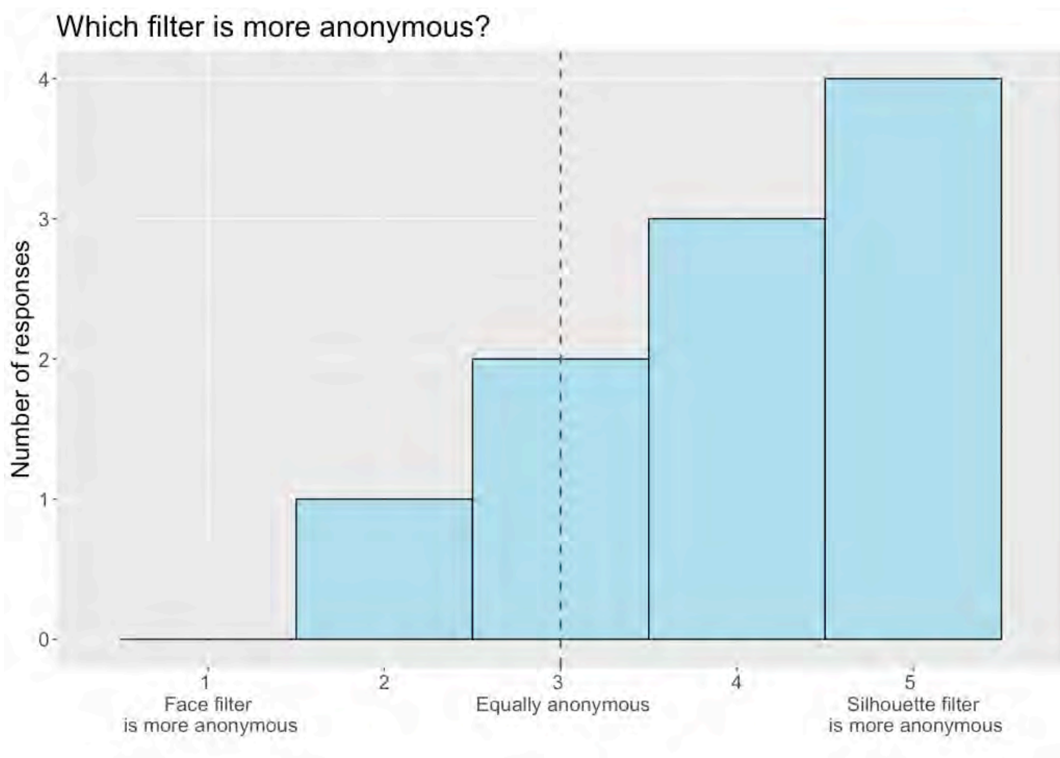


Fig. 4. Histogram of participant ratings of which filter is more anonymous (higher scores the indicate silhouette filter is more anonymous).

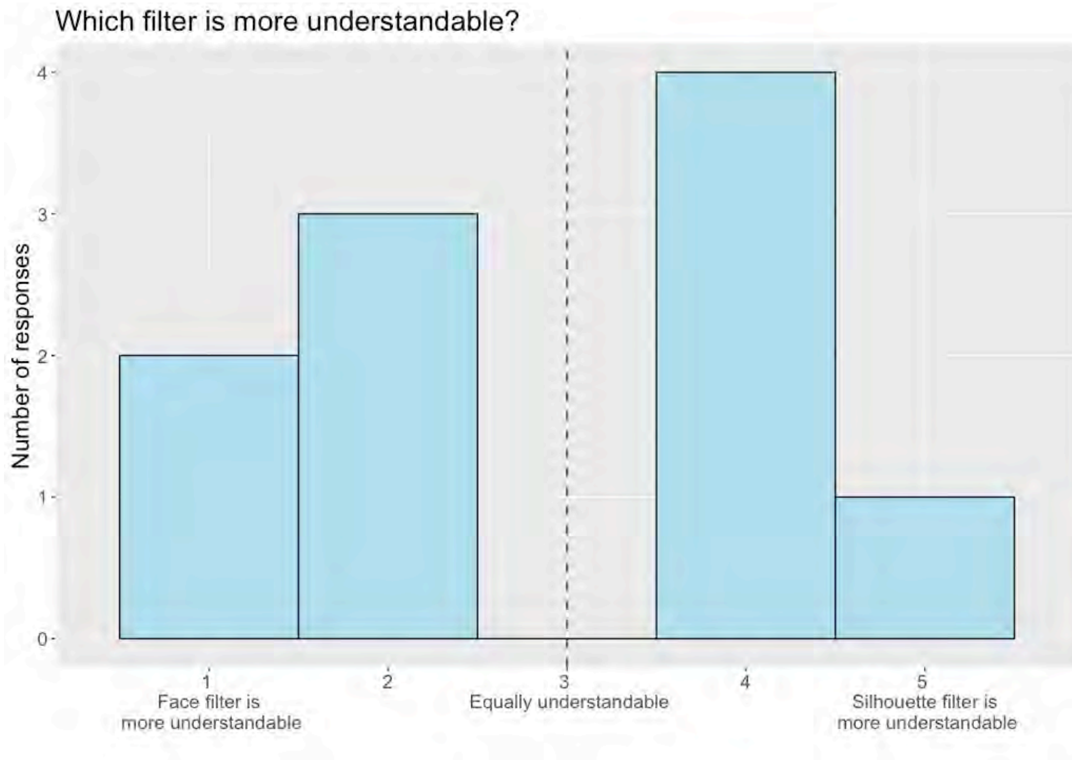


Fig. 5. Histogram of participant ratings of which filter is more understandable (higher scores the indicate silhouette filter is more understandable).

For some participants there was a trade-off between the two filters – the silhouette filter offered more privacy, but reduced comprehension. But for others, the silhouette filter was better on both counts: more private and easier to understand.

Feedback from Users

Participants liked the concept of the filters, though they raised several specific issues with the prototypes. For both the silhouette and the face mesh filter, there were several complaints about the face – one signer reports that the face mask ‘feels so robotic’ and doesn’t allow expressions to ‘fully come through’, and others label the face mask as ‘unnatural’ and ‘choppy’. For the green silhouette filter, one participant commented feeling ‘disconnected from the signer’ and states ‘facial movements were difficult to interpret’.

There was a mixed reaction to the hand skeleton in the silhouette filter. For some the hands were hard to see and follow, while others commented that it was surprising easy to understand the signing. Regardless, there was broad consensus on how to improve the filter: thicker lines and dots to distinguish the joints and changing the silhouette color (no neon green!).

Conclusion

Participants were intrigued by the possibilities afforded by these filters, particularly the silhouette filter. Users are mostly able to understand and recall the content and believed that the silhouette filter was effective at hiding the signer's identity. When comparing the two approaches we tested we also see a familiar tradeoff across filters between anonymity and ease of comprehension discussed by Lee et al. and Saunders et al., although the dot-and-line silhouette approach was surprisingly understandable. Several participants were intrigued to find that they did not mind the dot-and-line formation of the hands, and understood the silhouette signing as well as the signing in the face mesh filter, in which the hands were unobscured. While the results may seem unintuitive, as mentioned earlier work on sign perception also show deaf signers can recognize signs from moving dot displays (Poinzer et al., 1981; Leannah et al., 2022; Quandt et al., 2021), and fill in the gaps.

Nonetheless, in their feedback many participants expressed concerns about naturalness and ease of understandability for both filters. Like Lee et al. (2021), we found complaints about naturalness focused on the face mask. Despite using a detailed face estimation with over four hundred vertices, the mask still did not capture subtle facial movements precisely enough to satisfy viewers. Lee, Bragg, and others have suggested that naturalness could be achieved with increasingly human-like avatars. However, one of the original motivations for using MediaPipe is to identity a low resource approach that could work in real time on everyday devices. We want to keep to a lightweight silhouette model could also help minimize resources, compared to detailed deepfake avatars such as AnonySign (Saunders, et al., 2021).

In future work on silhouette-style filters, we suggest some simple changes could yield great improvements: a black or darkly colored silhouette, and on the hands, white dots and lines with increased thickness for visibility. The background can be occluded for increased anonymization. If users are turned off by unnatural faces, then we suggest reducing the salience of the face by reducing the number and/or size of vertices drawn, and foregrounding the hands.

The human mind is incredible at filling in gaps in perception and perhaps the most successful approach is one that allows the viewer to ‘connect the dots’ themselves, so to speak.

Acknowledgments

This work was supported by NSF #2150429 and #2131524.

Works Cited

Bragg, Danielle, et al. "Exploring collection of sign language datasets: Privacy, participation, and model performance." *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility*. 2020.

Herrmann, Annika, and Markus Steinbach, eds. Nonmanuals in sign language. Vol. 53. John Benjamins Publishing, 2013.

Leannah, Carly, Athena S. Willis, and Lorna C. Quandt. "Perceiving fingerspelling via point-light displays: The stimulus and the perceiver both matter." PloS one 17.8 (2022): e0272838.

Lee, Sooyeon, et al. "American Sign Language Video Anonymization to Support Online Participation of Deaf and Hard of Hearing Users." *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility*. 2021.

Lugaresi, Camillo, et al. "Mediapipe: A framework for building perception pipelines." *arXiv preprint arXiv:1906.08172* (2019).

Pfau, Roland, and Josep Quer. "Nonmanuals: Their Grammatical and Prosodic Roles." Sign Languages, edited by Diane Brentari, Cambridge University Press, Cambridge, 2010, pp. 381–402. Cambridge Language Surveys.

Poizner, Howard, Ursula Bellugi, and Venita Lutes-Driscoll. "Perception of American sign language in dynamic point-light displays." Journal of experimental psychology: Human perception and performance 7.2 (1981): 430.

Quandt, Lorna C., et al. "Enhanced biological motion perception in deaf native signers." Neuropsychologia 161 (2021): 107996.

Saunders, Ben, Necati Cihan Camgoz, and Richard Bowden. "AnonySign: Novel Human Appearance Synthesis for Sign Language Video Anonymisation." *2021 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021)*. IEEE, 2021.

Sisman, Berrak, et al. "An overview of voice conversion and its challenges: From statistical modeling to deep learning." IEEE/ACM Transactions on Audio, Speech, and Language Processing 29 (2020): 132-157.

Wilbur, Ronnie B. "Phonological and prosodic layering of nonmanuals in American Sign Language." The signs of language revisited. Psychology Press, 2013. 196-220.

Zhang, Fan, et al. "Mediapipe hands: On-device real-time hand tracking." *arXiv preprint arXiv:2006.10214* (2020). DOI: https://doi.org/10.48550/arXiv.2006.10214

Deaf and Hearing Small Group Inclusive Communication System

Yasmine Elglaly[1], Christa Miller[2], Chreston Miller[2], Rohan Patel[3], Spoorthy Annapareddy[3]
Washington State University[1], Virginia Tech[2], Rochester Institute of Technology[3]
elglaly@wwu.edu, millerch@vt.edu, chmille3@vt.edu, rp4437@rit.edu, sa3684@g.rit.edu

Abstract

Small groups of Deaf or hard-of-hearing (DHH) and hearing individuals find in-person communication challenging due to differences in preferred communication modalities (L. Elliot et al., Ntsongelwa and Rivera-Sánchez;). Groups of DHH and hearing individuals usually use systems with typing and speech recognition to communicate (Butler et al.; Glasser et al.; Mallory et al.; Stinson et al.; Marchetti et al.). However, these systems were often designed for asynchronous communication. To understand how design may address this gap, we conducted two studies. In study 1, we conducted semi-structured interviews and focus groups with 16 DHH and hearing participants. Study 1 informed the design of an inclusive group communication system, CollabAll. In study 2, we conducted a comparative study with 16 DHH and hearing participants to evaluate the benefits of CollabAll. Our empirical findings suggest that the ability to interject, quickly voice an opinion or challenge those who are holding the floor was a reoccurring communication need generally not available in existing group communication systems. The design of CollabAll facilitated interjecting using accessible buttons labeled with clear messages, e.g., Agree, Repeat, etc. The evaluation results indicated that group discussions are better structured and more efficient when texting is complemented with interjections support.

Key Words

Human-centered computing, Accessible technologies, small group, group communication, deaf, Deaf, assistive technology, accessibility, speaker diarization, interjection.

Introduction

Success in group work depends on effective and efficient communication (Leavitt). For groups of Deaf/Hard-of-Hearing (DHH) and hearing individuals, communication can be a challenge (Bauman and Murray; Bettencourt et al.; Ntsongelwa and Rivera-Sánchez; Sabila et al.) as many DHH individuals communicate in American Sign Language (ASL) as their primary language, but few hearing individuals know ASL (Mitchell et al.; Nakaji), or are familiar with Deaf Culture (Majocha et al.; Stinson and Liu). While ASL interpreters may facilitate group communication, interpreters are not always available (L. Elliot et al.) and difficult to find (Humphrey; Schwenke et al.). In the absence of interpreters, texting applications, e.g., Slack, and Speech-to-text (STT) applications, e.g., Microsoft Translate, are widely used to mediate group communication (Butler et al.; Glasser et al.; Mallory et al.; Stinson et al.; Marchetti et al.). However, these applications have limitations as they do not support DHH individuals in actively participating in discussions (Majocha et al.; Ntsongelwa and Rivera-Sánchez; L. Elliot et al.).

To mitigate these communication barriers, we developed an inclusive group communication system, called CollabAll, which involved two studies. In study 1, we employed semi-structured interviews and focus groups with 10 DHH individuals, 2 ASL interpreters, and 3 hearing individuals to inform our design. In study 2, we compared a mainstream texting system, Slack, to CollabAll with 4 groups of hearing and DHH participants. The findings of study 1 showed that DHH individuals find it difficult to take turns or interrupt in group communication with hearing individuals. Hence, CollabAll was designed to afford not only texting and STT, but also interjections, which came in the form of accessible buttons labeled with short messages, such as “Slow Down”, “Question”, and “Repeat”. The findings of study 2 indicate that CollabAll, and especially its interjecting modality, enable groups of mixed hearing ability to communicate more effectively. *The contribution of our work is multifold.* First, we present formative study findings on designing inclusive communication systems for groups of DHH and hearing individuals. Second, we present empirical findings on the usefulness of CollabAll. Last, we share the source code, the installation and deployment instructions of CollabAll (*CollabAll*; Kaushik).

Related Work

Prior research on DHH specific technology focused on one-way communication, such as capturing multiple sources of information during a lecture (Miller et al.) and real-time captions (Peng et al.). Little work was done on designing a digital solution to support active participation of DHH individuals in group discussions, however, the presence of an ASL interpreter was required (Peruma and Elglaly). Groups of DHH and hearing individuals usually communicate using text-based systems, e.g., word document and emails (L. Elliot et al.; Mallory et al.; L. B. Elliot et al.; Stinson et al.). However, DHH individuals were not satisfied with these systems due to lack of discussion support (L. Elliot et al.). STT, with its inaccuracy, was used to support DHH students in a hearing school to access the spoken lecture (Kheir and Way; Prietch et al.), and to support DHH workers in a hearing workplace (Mallory et al.).

Communication barriers lead to DHH group members feeling left out and excluded because it is difficult for them to put forth their opinions in a group discussion or take turns (Majocha et al.; Ntsongelwa and Rivera-Sánchez). A group conversation can manifest behaviors such as overlapping speakers and lapses (Schegloff). Hence, turn-taking organization is fundamental for effective and equitable communication (Schegloff). The length of the turn can be a word, a sentence or more (Sacks et al.), and can take the form of back channeling (feedback given by listener to show interest) (Ward), or an interruption (Gravano and Hirschberg). Turn-taking allocation among hearing individuals depends on auditory cues (Lala et al.; Scheetz), making it difficult for DHH individuals to take a turn (Hauser et al.; Meek). Hence, it is not surprising that DHH individuals tend to use text-based systems when interacting with hearing individuals, (L. Elliot et al.; Stinson et al.). Text-based communication systems, however, introduce two challenges: 1) Lack of simultaneous feedback, especially the non-verbal cues (Jokinen et al.); and 2) Disrupted turn adjacency, e.g., a response does not immediately follow a question (Anderson et al.; Herring). CollabAll was designed to facilitate turn-taking especially by using non-auditory backchanneling and interruptions affordances.

Study 1 - Formative Design

We designed CollabAll using a user-informed approach. This began with interviewing 4 DHH students and 2 ASL interpreters to learn about the challenges of group meetings. We then ran a semi-structured focus group with 9 individuals (DHH = 6 and hearing = 3) to obtain feedback on the design. The hearing participants had previous experience working with DHH individuals. Two out of the six DHH participants did not know ASL. The researchers met with the participants 2 times for 45-60 minutes with an ASL interpreter. The participants discussed their experiences and issues in group meetings when members have different levels of hearing ability. They brainstormed ideas, prototyped, and made design recommendations.

After coding the interviews, we analyzed comments that characterized experiences and communication barriers, if any. The DHH participants reported that they frequently experience situations where they communicate with hearing individuals without an interpreter present and relied on STT solutions and texting as an alternative communication method. They reported using some common tools such as Slack, Google docs, Notes, Microsoft Translate and Ava (“Notes”; *Ava - All-in-One Click Captions for All Conversations*; “Apps - Microsoft

Translator”). Of these only Ava is designed specifically to support DHH individuals in communicating with hearing people using STT, but DHH participants noted that the STT is not always accurate. Participants generally agreed that STT-based communication strategies may work in one-to-one conversation but lead to confusion in group meetings.

The participants envisioned a group meeting system that is inclusive for both DHH and hearing individuals and supports three key activities:

1. **Communication.** The participants considered typing the most common form of communication between DHH and hearing individuals.
2. **Catching up.** The participants suggested adding STT to the system to support individuals who are inclined to communicate using speech, and DHH individuals who often find it hard to follow spoken group discussions.
3. **Voicing an opinion.** The DHH participants wanted to be able to voice their opinion during a fast-paced discussion, e.g., through text-based interjections. They argued, however, about whether the interjections should be accompanied by audio feedback. Hearing participants thought it could be too distracting if the interjections were used frequently. The participants also suggested making the interjections editable, so each group could customize the interjections.

CollabAll was developed based on these findings to support small groups using three modalities, integrated and displayed in the same window, see Fig. 1. These modalities are:

1. **Typing.** Text messages are displayed in the chat window.
2. **Speaking.** Speech is transcribed and displayed as text. The microphone button has a micro interaction animation to show that speech is being recorded and transcribed.
3. **Interjecting.** Interjections are designed as large buttons, with a distinct label, color, and

icon. The interjections are displayed as the icon and text in the chat window.

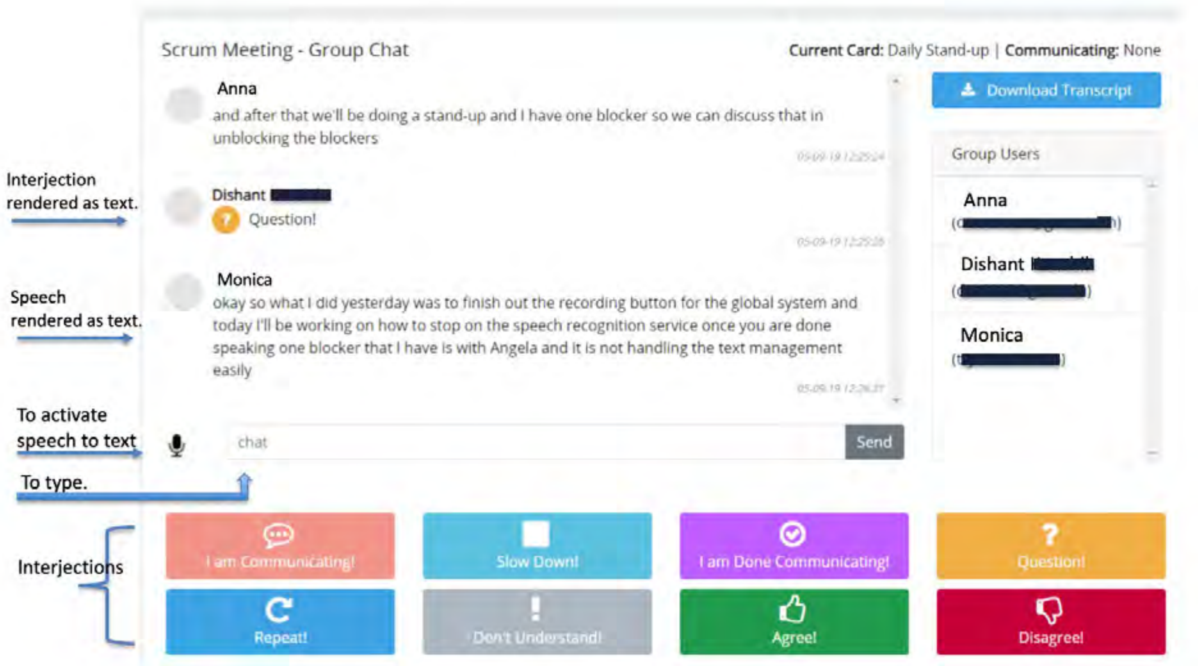


Fig. 1. User interface of CollabAll. The communication window displays the exchanged messages (left), users belong to the group, emails are hidden (right). At the bottom, there exists 3 communication methods: a microphone for speaking, text entry box for typing, and interjections buttons.

Typing and speaking modalities are meant to support communication, and interjecting is meant to support catching up and voicing an opinion. A group member can take a turn in discussions by typing or speaking. Group members can use interjections as a quick mechanism to convey various opinions, e.g., agreement, disagreement, etc., similar to the usage of vocal utterances or gestures in communication.

Software Development

CollabAll is built as a responsive web application using Express, Socket.io, Angular and MySQL. We used Web Speech API (Web Speech API - Web APIs | MDN) to render STT. The speech recognition service runs on the user's browser. To support in-person group conversations, we implemented speech diarization, which is the process of profiling speech based on the speaker (Anguera Miro et al.). The recognized speech is then displayed as text associated with its speaker profile. Similarly, the system displays the interjection's message, e.g. Repeat!, associated with the interjection initiator's profile. Both speech and interjections are sent to the server as chatting messages. The messages are displayed in the order they were received by the server. We implemented full-duplex communication so that users can send messages at the same time as STT interpretation. CollabAll is deployed on Huroku (*CollabAll*). The source code along with installation and deployment instructions are available on (Kaushik).

Discussion

We conducted a mixed-methods semi-experimental proof-of-concept study to examine CollabAll alongside Slack (Slack). The following research questions guided the study:

1. Which form of communication (text, interjections, speech) leads to better communication in small groups of mixed hearing ability collaborating in co-located contexts?
2. How can communication technology support turn-taking in groups of DHH and hearing individuals?

To answer these, we considered: group productivity (time to finish a task), turn taking (number of times each member took a turn), and active participation (volume of text by each member).

Method

We limited the size of the groups to 4 participants which is common in STEM education, (Kalaian et al.). We also chose to have 1 DHH participant and 3 hearing participants per group to mirror the prevailing experience of DHH individuals (L. B. Elliot et al.). The user study was structured so each group would complete a familiarization task followed by 3 tasks:

- Task 1: meeting time. We created 4 different schedules then asked participants to decide on a meeting time. There were only 4 possible periods where all members were available.

- Task 2: project ranking. We gave the participants 3 project ideas and asked them to rank them according to their own preference and come to a consensus.
- Task 3: role assignment. The participants were asked to decide on a role for each group member based on 4 options.

These tasks were designed to trigger communication, collaboration, and negotiation among the participants. The maximum duration of each task was set at 15 minutes. Each group used the systems in a different order to avoid ordering effects (see Table 1). In task 1, CollabAll instead of Slack was inadvertently provided to group 4 creating an imbalance. These data were excluded from the analysis.

Table 1. Participant distribution by group and what order the systems (Slack, CollabAll, and CollabAll+STT) were used by each group.

Group	Task 1 - calendar	Task 2 - project ideas	Task 3 - roles	Participants
Group 1	CollabAll	Slack	CollabAll+STT	PE1, PE2, PE3, PE4
Group 2	Slack	CollabAll+STT	Slack	PE5, PE6, PE7, PE8
Group 3	CollabAll	Slack	CollabAll+STT	PE9, PE10, PE11, PE12
Group 4	CollabAll	CollabAll+STT	Slack	PE13, PE14, PE15, PE16

Participants

Sixteen participants were recruited from Rochester Institute for Technology, 12 hearing students and 4 DHH students. Their age ranged from 21 to 29, 5 female (1 DHH) and 11 male (3 DHH). Participants’ primary languages were English (n = 9), Hindi (n = 3), ASL (n = 2), Telugu (n = 1), and Gujarati (n = 1). All participants had used Slack. They were compensated $20.

Environment

The study environment is shown in Fig. 2. Each participant was given a laptop and a lapel microphone. The sensitivity of the microphone was adjusted to avoid duplication of speech recognition. Two cameras were located on opposite corners of the room for recording.

Fig. 2. A screenshot for one of the user study sessions. The four group members sit together around a table as in regular meetings. On the left side a researcher taking notes.

Analysis

To analyze the messages sent via each system, we based our analysis on character count, which included spaces, punctuation, reactions and emojis. For our analysis, an emoji, a reaction and an interjection were each counted as one "character". Looking at total characters, the DHH participants averaged 581 characters per person ($\sigma = 20$). Whereas hearing participants averaged 695 characters per person ($\sigma = 3$). The results (Fig. 3) indicated that communication in CollabAll was slightly more efficient. Participation was then examined based on the number of turns taken and organized by system and participant type (Figure 4). The biggest number of taking turns occurred with CollabAll+STT (n=20), followed by Slack (n=15), then CollabAll (n = 9).

Table 2. The weighted mean of turns taken grouped by system and participant hearing ability where a turn consisted of text message sent, reaction sent, or interjection sent.

System	DHH	Hearing
Slack	8	7
CollabAll	5	4
CollabAll+STT	9	11

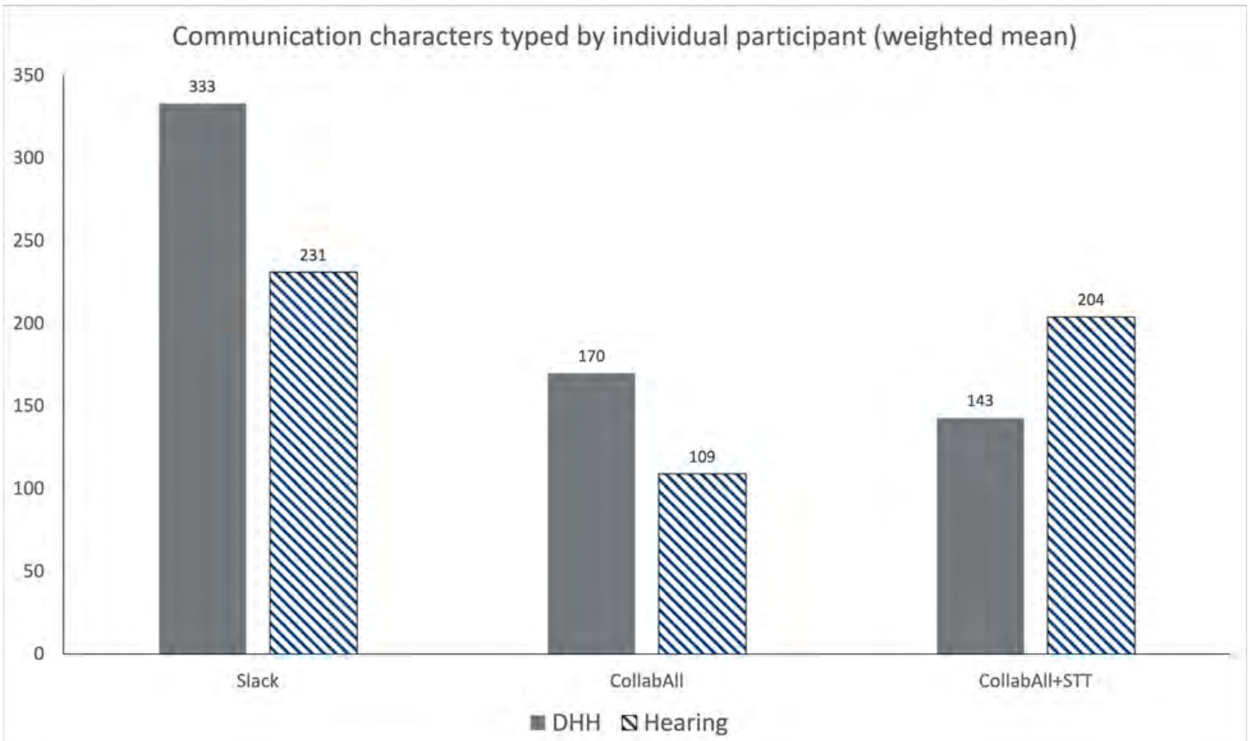


Fig. 3. Bar chart comparison of characters communicated per individual grouped by system and participant hearing ability. This includes characters typed, characters spoken, spaces, punctuation, emojis, reactions, and interjections.

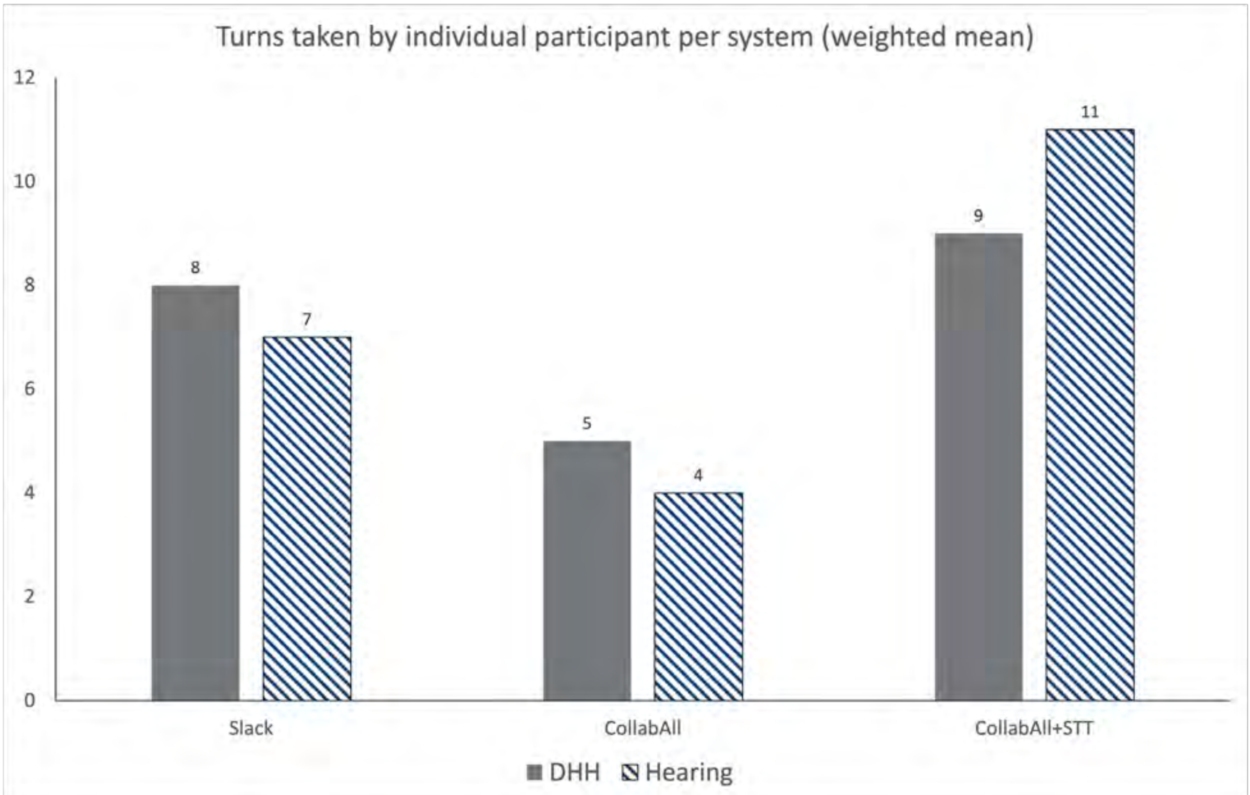


Fig. 4. Bar chart comparison of turns taken per individual grouped by system and hearing status.

Another area of participation analysis was on the use of emojis (inline and reactions) and interjections. Emojis were available in Slack only and were used a total of 6 times across all tasks. This is a stark contrast to the use of interjections. The interjections feature of CollabAll and CollabAll+STT was used a total of 52 times (CollabAll = 6 and CollabAll+STT = 46) see Fig. 5. All participants used the interjection feature at least once. DHH individuals used the interjection feature an average of 4 times per person and a total of 14 times. In comparison, hearing (H) individuals used the interjection feature an average of 3 times per person and 38 times in total. The interjection “Agree” was used the most at a total of 26 times by 15 participants (*H* = 12, and *DHH* = 3).

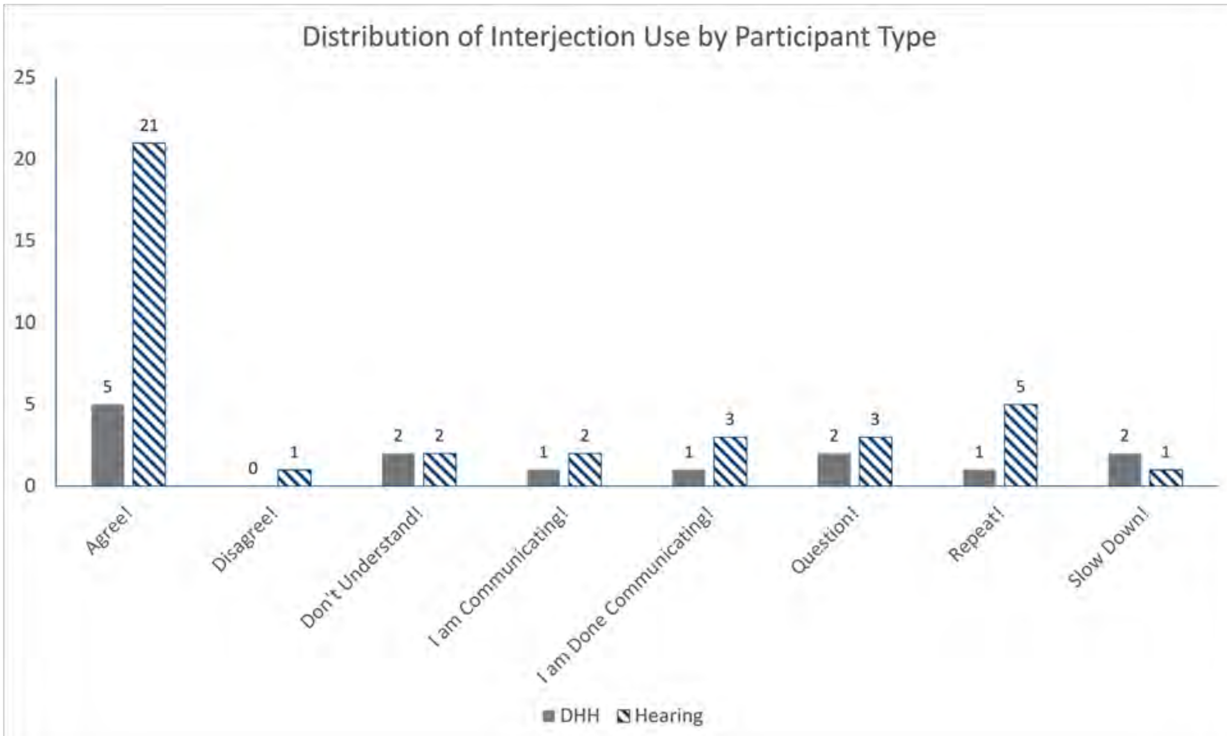


Fig. 5. The number of times each interjection was used as categorized by hearing status.

The last element of participation analysis was to consider the time to complete tasks. The results appear in Table 3. In general, CollabAll groups reached consensus more quickly.

Table 3. Completion time based on the system type for Slack, CollabAll, CollabAll+STT.

System	mean completion time (s)	Deviation σ
Slack	405	0.00219
CollabAll	217	0.00087
CollabAll+STT	389	0.00085

The video of each session was annotated by two annotators who were members of the research team using the VIA video annotation tool (Dutta and Zisserman). The annotations were made using a defined dictionary (Guest and MacQueen). Using Cohen's kappa we found fair intercoder agreement $\kappa = 0.23$. The annotations showed that the largest portion of collaborative communication outside the system was speaking $729s$ (16%). Cross-referencing this to the sessions where CollabAll+STT was used, we determined that 329s were use of the STT feature. Meaning that 45% of the speaking annotations are actually a use of the system. The second most frequent annotation was gestures 116s or 2.5%. From this we conclude that most communication happened within the system.

The qualitative analysis indicated that participants tended to bounce between topics when they used Slack but stayed more on topic using CollabAll. For example, one group of participants noticeably bounced from topic to topic using Slack but when using CollabAll, they tended to discuss one topic at a time, and then move to the next.

Based on our analysis, we found that with the texting only modality (Slack condition), participants took more time to complete their tasks. We also found that using CollabAll groups tended to reach consensus more quickly, possibly connected to the availability of interjections. Last, we found that CollabAll groups tended to stay on topic better.

Conclusions

DHH individuals face the challenge of becoming part of the “flow” of the conversation as they communicate in-person with hearing individuals. CollabAll assists in this respect as it allows the users to provide “cues” (interjections) that communicate their opinion, in addition to texting and STT features that accommodate the communication preferences for the DHH and the hearing individuals. Our study showed how group productivity, turn taking, and active participation can be facilitated through a combination of text messaging, interjections, and STT. It is notable that CollabAll generally allowed for greater turn taking. Along the same lines, the analysis indicates that text alone supports fairly equal communication for groups of mixed ability. The user study further showed regular use of the interjections and STT features by both hearing and DHH individuals.

The number of DHH participants in our combined studies was relatively small (14 in total). The participants were students in a university with a significant number of DHH students. Different group dynamics can occur if the hearing participants were less familiar with Deaf culture. As CollabAll was evaluated in a laboratory study, different results may occur if it was used in the wild. In future work, we plan on evaluating CollabAll in-the-wild to further understand the effectiveness of the system and learn more about dynamics of groups with mixed hearing ability.

Works Cited

Anguera Miro, Xavier, et al. "Speaker Diarization: A Review of Recent Research." *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 2, Feb. 2012, pp. 356–70. *DOI.org (Crossref)*, https://doi.org/10.1109/TASL.2011.2125954.

"Apps - Microsoft Translator." *Microsoft Translator for Consumers*, https://www.microsoft.com/en-us/translator/apps/. Accessed 3 Oct. 2022.

Ava - All-in-One Click Captions for All Conversations. https://ava.me/. Accessed 3 Oct. 2022.

Bauman, H.-Dirksen L., and Joseph J. Murray. *Deaf Gain: Raising the Stakes for Human Diversity*. U of Minnesota Press, 2014.

Bettencourt, Genia M., et al. "Disability in Postsecondary STEM Learning Environments: What Faculty Focus Groups Reveal About Definitions and Obstacles to Effective Support." *Journal of Postsecondary Education and Disability*, vol. 31, no. 4, Jan. 2019, pp. 383–96.

Butler, Janine, et al. "Exploration of Automatic Speech Recognition for Deaf and Hard of Hearing Students in Higher Education Classes." *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*, Association for Computing Machinery, 2019, pp. 32–42. *ACM Digital Library*, https://doi.org/10.1145/3308561.3353772 .

CollabAll. https://collaballsystem.herokuapp.com/login. Accessed 10 Oct. 2022.

Dutta, Abhishek, and Andrew Zisserman. "The VIA Annotation Software for Images, Audio and Video." *Proceedings of the 27th ACM International Conference on Multimedia*, ACM, 2019, pp. 2276–79. *DOI.org (Crossref)*, https://doi.org/10.1145/3343031.3350535.

Elliot, Lisa, et al. “Deaf and Hard of Hearing Individuals’ Perceptions of Communication with Hearing Colleagues in Small Groups.” Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility, ACM, 2016, pp. 271–72. ACM Digital Library, https://doi.org/10.1145/2982142.2982198.

Elliot, Lisa B., et al. “College Students’ Perceptions of the C-Print Speech-to-Text Transcription System.” *The Journal of Deaf Studies and Deaf Education*, vol. 6, no. 4, Oct. 2001, pp. 285–98. *academic.oup.com*, https://doi.org/10.1093/deafed/6.4.285.

Glasser, Abraham, et al. “Deaf, Hard of Hearing, and Hearing Perspectives on Using Automatic Speech Recognition in Conversation.” *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, ACM, 2017, pp. 427–32. *ACM Digital Library*, https://doi.org/10.1145/3132525.3134781.

Gravano, Agustín, and Julia Hirschberg. “Turn-Taking Cues in Task-Oriented Dialogue.” *Computer Speech & Language*, vol. 25, no. 3, July 2011, pp. 601–34. *DOI.org (Crossref)*, https://doi.org/10.1016/j.csl.2010.10.003.

Guest, Greg, and Kathleen M. MacQueen. *Handbook for Team-Based Qualitative Research*. Rowman Altamira, 2008.

Hauser, Peter C., et al. “Deaf Epistemology: Deafhood and Deafness.” *American Annals of the Deaf*, vol. 154, no. 5, 2010, pp. 486–92.

Herring, S. C. “Interactional Coherence in CMC.” *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. 1999. HICSS-32. Abstracts and CD-ROM of Full Papers*, IEEE Comput. Soc, 1999, p. 13. *DOI.org (Crossref)*, https://doi.org/10.1109/HICSS.1999.772674.

Humphrey, Carrie. *Job Satisfaction, Role Strain, Burnout, and Self-Care among American Sign Language/English Interpreters*. Western Oregon University, Dec. 2015. *Zotero*, https://digitalcommons.wou.edu/cgi/viewcontent.cgi?article=1030&context=theses.

Jokinen, Kristiina, et al. "Gaze and Turn-Taking Behavior in Casual Conversational Interactions." *ACM Transactions on Interactive Intelligent Systems*, vol. 3, no. 2, July 2013, pp. 1–30. *DOI.org (Crossref)*, https://doi.org/10.1145/2499474.2499481.

Kalaian, Sema A., et al. "Effectiveness of Small-Group Learning Pedagogies in Engineering and Technology Education: A Meta-Analysis." *Journal of Technology Education*, vol. 29, no. 2, 2018, pp. 20–35.

Kaushik, Dishant. *CollabAll*. 2019. 12 Sept. 2020. *GitHub*, https://github.com/dkaushik95/CollabAll.

Kheir, Richard, and Thomas Way. "Inclusion of Deaf Students in Computer Science Classes Using Real-Time Speech Transcription." *Proceedings of the 12th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education - ITiCSE '07*, ACM Press, 2007, p. 261. *DOI.org (Crossref)*, https://doi.org/10.1145/1268784.1268860.

Lala, Divesh, et al. "Attentive Listening System with Backchanneling, Response Generation and Flexible Turn-Taking." *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, Association for Computational Linguistics, 2017, pp. 127–36. *DOI.org (Crossref)*, https://doi.org/10.18653/v1/W17-5516.

Leavitt, Harold J. "Some Effects of Certain Communication Patterns on Group Performance." *The Journal of Abnormal and Social Psychology*, vol. 46, 1951, pp. 38–50. *APA PsycNet*, https://doi.org/10.1037/h0057189.

Majocha, Megan, et al. "'Everyone Was Nice…But I Was Still Left Out': An Interview Study About Deaf Interns' Research Experiences in STEM †." *Journal of Microbiology & Biology Education*, vol. 19, no. 1, Jan. 2018. *Crossref*, https://doi.org/10.1128/jmbe.v19i1.1381.

Mallory, James R., et al. "Personal Perspectives on Using Automatic Speech Recognition to Facilitate Communication Between Deaf Students and Hearing Customers." *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, ACM, 2017, pp. 419–21. *ACM Digital Library*, https://doi.org/10.1145/3132525.3134779.

Marchetti, Carol, et al. "Faculty Perspectives on Developing Strategies to Improve Access in Diverse Postsecondary Classrooms." *Pedagogies*, Jan. 2019.

Meek, David R. "Dinner Table Syndrome: A Phenomenological Study of Deaf Individuals' Experiences with Inaccessible Communication." *The Qualitative Report*, vol. 25, no. 6, June 2020, pp. 1676-1694,1676A.

Miller, Dorian, et al. "Semi-Transparent Video Interfaces to Assist Deaf Persons in Meetings." *Proceedings of the 45th Annual Southeast Regional Conference*, Association for Computing Machinery, 2007, pp. 501–06. *ACM Digital Library*, https://doi.org/10.1145/1233341.1233431.

Mitchell, Ross E., et al. "How Many People Use ASL in the United States? Why Estimates Need Updating." *Sign Language Studies*, vol. 6, no. 3, 2006, pp. 306–35. *DOI.org (Crossref)*, https://doi.org/10.1353/sls.2006.0019.

Nakaji, Melanie C. A., *Dramatic Turnaround in a Classroom of Deaf Students*. p. 8.

"Notes." *App Store*, https://apps.apple.com/us/app/notes/id1110145109. Accessed 3 Oct. 2022.

Ntsongelwa, Avela, and Milagros Rivera-Sánchez. *Intercultural Communication between Deaf and Hearing Students at a South African University*. Dec. 2018, https://doi.org/10.18820/24150525/Comm.v23.11.

Peng, Yi-Hao, et al. “SpeechBubbles: Enhancing Captioning Experiences for Deaf and Hard-of-Hearing People in Group Conversations.” *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, ACM, 2018, pp. 1–10. *DOI.org (Crossref)*, https://doi.org/10.1145/3173574.3173867.

Peruma, Anthony, and Yasmine N. Elglaly. “CollabAll: Inclusive Discussion Support System For Deaf and Hearing Students.” *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, ACM, 2017, pp. 315–16. *ACM Digital Library*, https://doi.org/10.1145/3132525.3134800.

Prietch, Soraia Silva, et al. “Application Requirements for Deaf Students to Use in Inclusive Classrooms.” *Proceedings of the Latin American Conference on Human Computer Interaction*, ACM, 2015, pp. 1–8. *DOI.org (Crossref)*, https://doi.org/10.1145/2824893.2824898.

Sabila, Paul S., et al. “Chapter 14 - FORWARD Workshops: Strategies for Inclusion of the Deaf and Hard of Hearing.” *Forward to Professorship in STEM*, edited by Rachelle S. Heller et al., Academic Press, 2016, pp. 223–43. *ScienceDirect*, https://doi.org/10.1016/B978-0-12-800855-3.00014-3.

Sacks, Harvey, et al. "A Simplest Systematics for the Organization of Turn Taking for Conversation**This Chapter Is a Variant Version of 'A Simplest Systematics for the Organization of Turn-Taking for Conversation,' Which Was Printed in Language, 50, 4 (1974), Pp. 696–735. An Earlier Version of This Paper Was Presented at the Conference on 'Sociology of Language and Theory of Speech Acts,' Held at the Centre for Interdisciplinary Research of the University of Bielefeld, Germany. We Thank Dr. Anita Pomerantz and Mr. Richard Faumann for Pointing out to Us a Number of Errors in the Text." *Studies in the Organization of Conversational Interaction*, Elsevier, 1978, pp. 7–55. *DOI.org (Crossref)*, https://doi.org/10.1016/B978-0-12-623550-0.50008-2.

Scheetz, Nanci A. *Deaf Education in the 21st Century: Topics and Trends*. Pearson Higher Ed, 2011.

Schegloff, Emanuel A. "Overlapping Talk and the Organization of Turn-Taking for Conversation." *Language in Society*, vol. 29, no. 1, Jan. 2000, pp. 1–63. *DOI.org (Crossref)*, https://doi.org/10.1017/S0047404500001019.

Schwenke, Tomina J., et al. "Sign Language Interpreters and Burnout: The Effects of Perfectionism, Perceived Stress, and Coping Resources." *Interpreting*, vol. 16, no. 2, 2014, pp. 209–32.

Slack. "Slack Is Your Digital HQ." *Slack*, https://slack.com/. Accessed 3 Oct. 2022.

Stinson, M., and Y. Liu. "Participation of Deaf and Hard-of-Hearing Students in Classes with Hearing Students." *The Journal of Deaf Studies and Deaf Education*, vol. 4, no. 3, July 1999, pp. 191–202, https://doi.org/10.1093/deafed/4.3.191.

Stinson, Michael, et al. "Using Automatic Speech Recognition to Facilitate Communication Between an Individual Who Is Hearing and One Who Is Deaf or Hard of Hearing." *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, ACM, 2017, pp. 407–08. *ACM Digital Library*, https://doi.org/10.1145/3132525.3134797.

Ward, Nigel G. *Prosodic Patterns in English Conversation*. Cambridge University Press, 2019.

Web Speech API - Web APIs | MDN. https://developer.mozilla.org/en-US/docs/Web/API/Web_Speech_API. Accessed 3 Oct. 2022.

The Journal on Technology and Persons with Disabilities

VR Training to Facilitate Blind Photography for Navigation

Jonggi Hong, James M. Coughlan
Smith-Kettlewell Eye Research Institute
jonggi.hong@ski.org, coughlan@ski.org

Abstract

Smartphone-based navigation apps allow blind and visually impaired (BVI) people to take images or videos to complete various tasks such as determining a user's location, recognizing objects, and detecting obstacles. The quality of the images and videos significantly affects the performance of these systems, but manipulating a camera to capture clear images with proper framing is a challenging task for BVI users. This research explores the interactions between a camera and BVI users in assistive navigation systems through interviews with BVI participants. We identified the form factors, applications, and challenges in using camera-based navigation systems and designed an interactive training app to improve BVI users' skills in using a camera for navigation. In this paper, we describe a novel virtual environment of the training app and report the preliminary results of a user study with BVI participants.

Keywords

Blind photography; virtual reality; walk light; navigation; blindness and low vision.

Introduction

Acquiring skills to take photos or videos with good quality has become a necessity for blind and low vision (BVI) people as they actively use cameras for various tasks such as saving memories and performing text recognition (Jayant *et al.* 2011). Using assistive mobile apps is also an essential reason for BVI individuals to use a camera. Smartphone-based navigation apps such as NaviLens and Clew enable BVI people to navigate indoors and outdoors independently by providing directional guidance based on images or videos of BVI users' surroundings. OrCam and Envision Glasses, which read a text and recognize objects, provide functions that detect traffic signs or landmarks for BVI users to navigate outdoors independently. Prior studies in computer vision and assistive technology developed camera-based navigation systems for BVI people to detect landmarks (Serrão *et al.*, 2012), objects (Afif *et al.*, 2020), and obstacles (Tapu *et al.*, 2013) or communication systems between BVI users and sighted people who interact with the BVI users through images or videos (Bigham *et al.* 2010).

However, taking photos or videos with good quality is challenging for BVI people due to the difficulty of image framing and focusing (Bigham *et al.* 2010). Prior studies employed non-visual guidance to help BVI individuals manipulate a camera better (Jayant *et al.* 2011, Vazquez *et al.* 2012, Manduchi *et al.* 2014, Lee *et al.* 2019). These interfaces detect an object of interest in users' images with computer vision techniques and provide guidance based on the object's position for better image framing. However, computer vision techniques are still error-prone for various reasons such as mismatches between training and real-world data, quality of images, and challenging light conditions. The audio and haptic guidance can be hard to perceive, distracting, or may raise privacy concerns (Easwara *et al.* 2015). To overcome these limitations, we aim to enable BVI users to learn how to take clear, properly framed images without audio and haptic guidance through interactive training in virtual reality (VR).

As VR holds a great opportunity to allow BVI individuals to explore simulated areas safely with augmented haptic and audio feedback, prior studies leveraged VR to enable BVI people to explore unfamiliar places in a virtual space (*e.g.*, Kunz *et al.*, 2018). To simulate navigation with VR effectively, Kreimeier *et al.* (Kreimeier *et al.*, 2020) evaluated different form factors for virtual navigation such as treadmills and trackers on the ankles. While these approaches employed VR to enable BVI users to have the experience of navigating unfamiliar places, we aim to improve BVI users' skills to interact with a camera for navigation. Based on BVI people's interactions with a camera for navigation identified through interviews, we designed an interactive training system using VR where BVI individuals can try out a simulated walk-light detector in a safe indoor environment with feedback to correct their gestures that may cause poor-quality images. The user study showed that interactive training allows BVI people to manipulate a camera properly when they scan their environments to find a walk light.

Discussion

As a preliminary study, we interviewed BVI participants to draw from their experiences with a camera for navigation when designing an interactive training tool. We report on the intermediate results from a user study that includes a training session followed by a task of using a walk light detector in a real environment through a user study.

Method

We conducted interviews with BVI individuals and used thematic coding to identify the main themes in their responses (Barun and Clarke, 2006). We recruited 10 participants (6 female and 4 male) from our email lists, whose ages ranged from 30 to 75 (*M*=46.4, *SD*=18.3). Seven participants reported being totally blind, two reported having light perception, and one reported being legally blind. Only two participants ever had full vision. Seven participants were congenitally blind. They have had the current level of vision for 35.5 years on average (*SD*=21.4). Nine participants had smartphones and used them for 10.7 years on average (*SD*=4.5). One participant never used a smartphone. All participants have used a camera. Seven of them reported using a camera several times a week or more (Figure 1a).

The interview includes topics relevant to BVI people's experience in using a camera and camera-based navigation systems: the methods to learn to take photos or videos; frequency of using assistive navigation systems; gestures to use a camera for assistive navigation systems; frequency of using a camera for navigation; apps or tools participants used with a camera for navigation; challenges of using a camera for navigation.

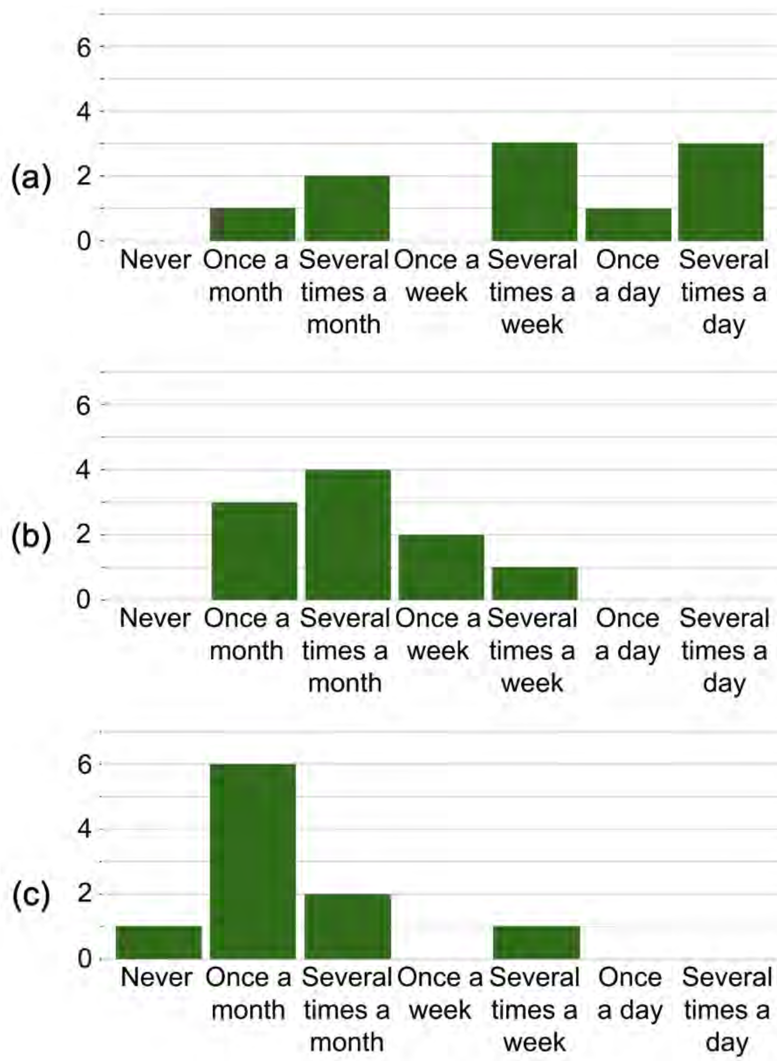


Fig. 1. Blind participants' responses to questions about (a) frequency of using a camera, (b) frequency of using navigation systems, and (c) frequency of using a camera for navigation systems.

Findings

The most common tasks where the participants used a camera were reading text (N=8), keeping or sharing memories (N=6), video calling (N=5), and identifying objects (N=3). We found only four participants learned or practiced using a camera for these tasks. Two of the participants practiced with the feedback from assistive apps (*e.g.*, sound feedback to indicate the position of a document in the camera frame, whether the app recognizes something or not). The other two participants got feedback from sighted people to learn to take good-quality photos.

All participants reported using navigation systems (both with and without a camera) at least once a month (Figure 1b). They used sensor-based (GPS, compass) navigation systems such as Google Maps and BlindSquare (N=10), apps for video calls with sighted people such as Aira and BeMyEyes (N=7), and computer-vision systems such as OneStep Reader and Google Lookout (N=2). Most participants reported using a camera for navigation once a month (N=6) as shown in Figure 1c. While using a camera, the participants aimed a camera at a target object by following directions from a sighted person in a video call (N=7), maintaining a certain camera height or orientation (N=4), moving the camera around to find a target object (N=1), and touching a target object (N=1). When using a camera for navigation, participants had challenges in image framing (N=7), controlling network connection (N=4), adjusting light conditions (N=3), focusing (N=2), and holding the camera steadily (N=1).

When asked what participants wanted to capture with a camera for navigation, most of them (N=7) wanted to capture landmarks (*e.g.*, stores, restaurants). Other responses were door (*i.e.*, room, N=5), person (N=5), sign (N=5), and walk light or traffic light (N=3). The preferred form factors of a camera for navigation were smart glasses (N=8), a smartphone (N=3), and a shoulder-mounted camera (N=1). The difficulty of aiming a camera affects participants' preference in the form factors. P4 who preferred smart glasses mentioned *"No need to point the camera (on glasses) to a specific place."* On the other hand, P8 preferred a smartphone because the participant did not want to carry multiple devices, saying *"[...] I don't like glasses because it is cumbersome to have glasses and masks together."*

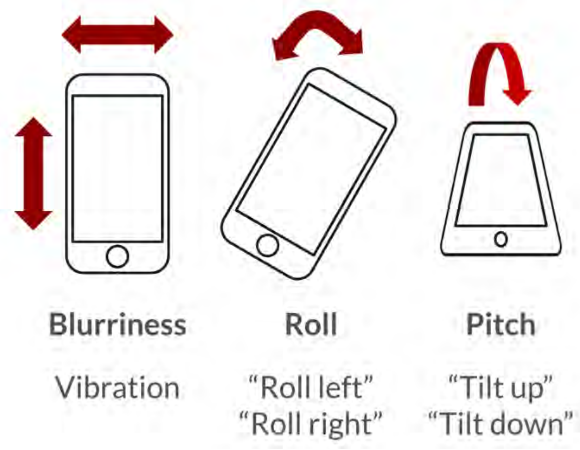


Fig. 2. Descriptions of audio and haptic feedback in the interactive training app.

Interactive Training App for Using a Walk-Light Detector

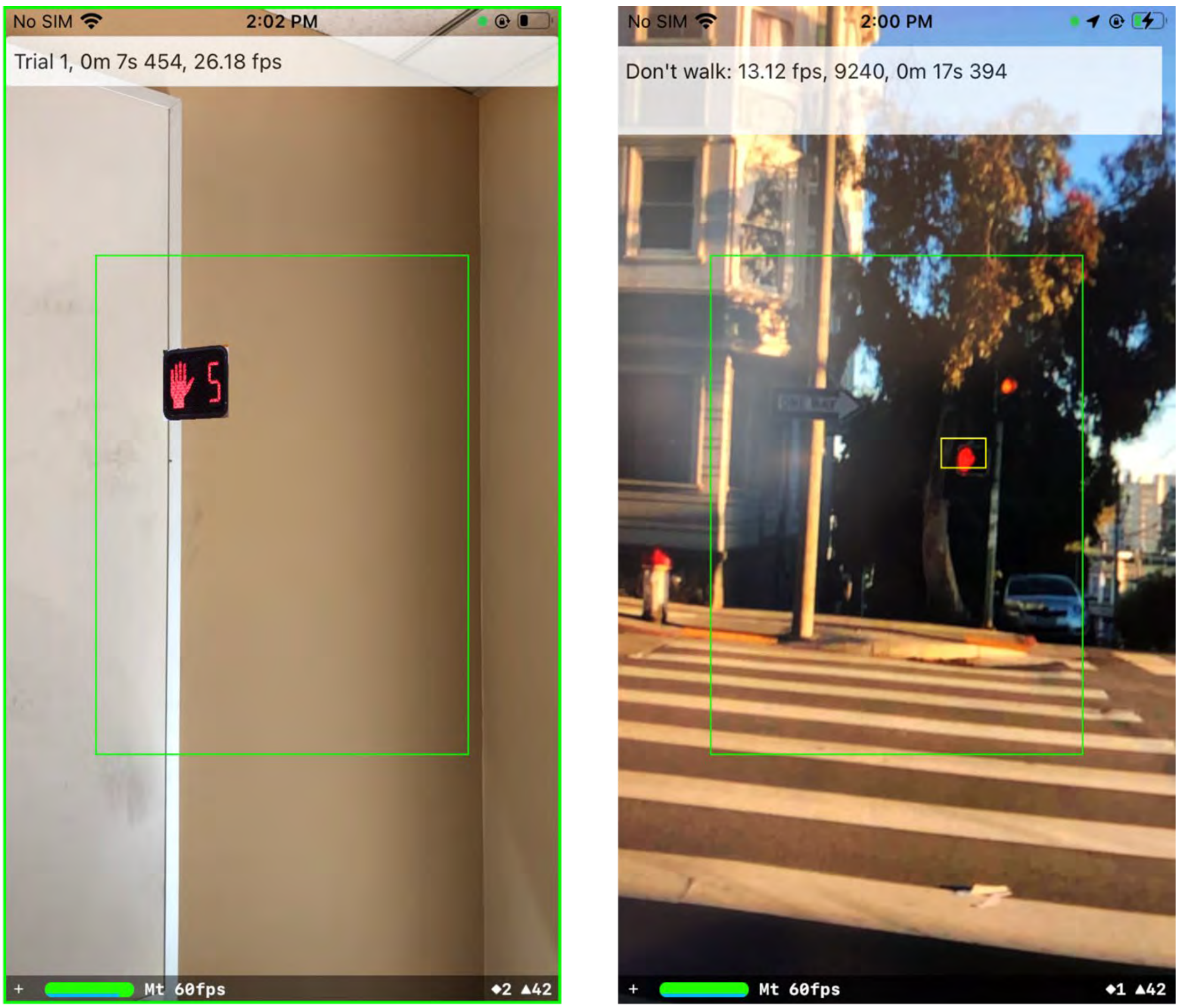
Based on the findings in the interview, we designed and implemented an interactive training app that helps BVI individuals get used to interactions with a camera for navigation. While the interview revealed that BVI people are interested in recognizing various objects using their cameras (e.g., door, person, sign, walk light), in this study, we focused on detecting a walk light as a case of using a camera for navigation. Our BVI participants also mentioned aiming a camera at a target object and taking clear photos as the main challenges in using camera-based assistive technology. Therefore, we sought to improve BVI users' ability to aim a camera at a walk light and take clear photos. We chose the smartphone as a form factor because it is a preferred form factor based on our interview results and because it is the most commonly used device to use a camera with assistive technology.

We built an interactive training app where BVI individuals practice using a camera for a walk-light detector at a virtual traffic intersection. We implemented the virtual environment using ARKit in iOS on an iPhone 8 as shown in Figure 3a. BVI users can interact with the virtual environment with smartphones without additional devices such as head-mounted displays or wearable devices. The app provides audio and haptic feedback to BVI users related to two interactions, image framing and focusing, while they use a walk light detector that indicates the status of the walk light (Walk, Don't walk, Countdown) in real-time. The virtual walk lights have the same size, distance, and height as walk lights in the real world under typical viewing conditions. The size and heights of the virtual walk lights were chosen based on the design guidelines for pedestrian control features from the US Department of Transportation (https://mutcd.fhwa.dot.gov/htm/2009/part4/part4e.htm). The distances to the virtual walk lights were based on the widths of roads around our institution, ranging from 12.47 to 31.24 meters.

This training app is designed to simulate some of the experience of using a real walk light detector app (which we created to function at real traffic intersections), except that the training app's feedback on proper camera orientation is something that is missing from the real walk light detector app. The main function of the training app is to indicate when the virtual walk light is visible to the camera and what state the walk light is in. To allow BVI users to practice taking a clear photo or video for a walk light detector, the app provides audio and haptic feedback as shown in Figure 2. The feedback includes vibration to indicate overly fast camera movement that may cause blurriness of images. The app vibrates when a walk light moves farther than five pixels between two consecutive frames in the camera video stream as we observed that the performance of the walk light detector degrades under this condition due to image blur. It also provides verbal warnings for BVI users to avoid tilting the phone up or down too far from the horizon and rolling the phone to the left or right.

Why a Virtual Environment?

Virtual reality has characteristics that enable the creation of an accessible training environment. It allows BVI users to try a navigation system in a safe environment (*e.g.*, a quiet room) before trying it in the real world (*e.g.*, a real traffic intersection). It allows them to have fewer distractions that are likely to occur in the real world (*e.g.*, pedestrians, noises). It also provides our training app with full control of the environment as the app can track the positions of virtual objects and the camera. Last, this approach allows BVI people to access the training environment independently with little help from sighted people because the training environment works in an empty space without installing any physical materials (*e.g.*, markers, beacons) in their environment.



(a) (b)

Fig. 3. Screenshots of (a) the interactive training app and (b) the walk light detector app. The green rectangles in the apps indicate the center area.

Walk Light Detector for a User Study

To evaluate the effect of using our interactive training app, we will conduct a user study where BVI people will first have a practice session with our interactive training app and will then use the walk light detector in a real environment. For the user study, we implemented a walk light detector using an off-the-shelf object detector, YOLOv2 (Redmon *et al*., 2016) object detection model. We fine-tuned the object detection model using transfer learning with photos of traffic intersections near our institution. A researcher in our team labeled the photos with bounding boxes around walk lights with three different classes: Walk, Count down, and Don't walk. Examples of the photos are shown in Figure 4.

A challenge of detecting a walk light is that the target object appears small under typical viewing conditions (the walk light is about 12-32 meters away in our study), which makes it hard for the object detection model to detect walk lights. To address this challenge, we designed the app to crop the center area of the image as input for the walk light detector and provide feedback about the state of a walk light when it is in the area as shown in Figure 3b.

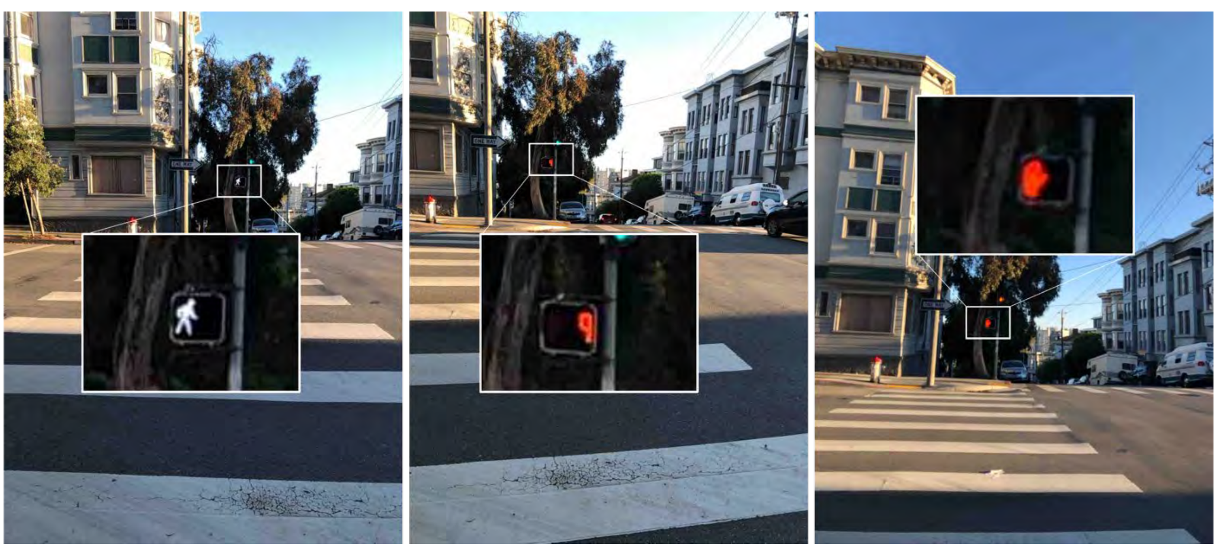


Fig. 4. Samples of the photos used to train a walk light detector model. Photos with “Walk”, “Count down”, and “Don't walk” signals from left to right.

Results from a Pilot Study with BVI Participants

Before conducting a user study, we piloted a preliminary version of it with two BVI participants. The participants first completed a training session (one used the interactive VR training, and the other listened to verbal instructions with a demo video of the walk light detector). Afterward, both participants then used the walk light detector app outdoors. The main findings of this pilot study were to confirm that (a) the participant who received VR training was able, in multiple trials, to successfully locate the desired walk light (which was located at a 90-degree angle relative to a “distractor” walk light located in a different direction) and to indicate when the light entered the Walk state; and (b) both participants were able to complete the real-world version of this task in an outdoor setting, in which they walked with the experimenters along a path with eight walk lights. Based on our experience with the pilot study, we made a few changes to the user study: (a) since some walk lights are accompanied by accessible pedestrian signals (APS), which provide real-time audio cues about the walk light status, we included different types of intersections, including those with and without walk lights, and walk lights with and without APS; we asked the participants to ascertain if there is a walk light in each trial before providing information about it if it is present; (b) we improved the performance of the walk light detector to decrease the incidence of misrecognition errors.

Evaluating VR Training with a User Study

We increased the number of training images in order to increase the test accuracy of the walk light detector to 98.1%. During the pilot studies, we observed that BVI participants assumed that the intersections have walk lights, though determining whether a walk light exists or not at an intersection is a challenge for BVI users in practice. To reflect the practical usage scenario, we included intersections with no walk light. With these modifications, we conducted a user study with seven BVI participants (4 female and 3 male). Six of the participants ranged in age from 39 to 73 (*M*=38, *SD*=12.7); a participant in her 30s declined to provide her exact age. We randomly assigned the three and four participants to two groups: VR training (VR) and verbal instructions (VI) groups, respectively. All but one participant in each group completed all trials successfully (i.e., detected the Walk state of the walk lights or distinguished intersections with no walk light). A participant (P5, VR) could not capture a walk light at one intersection. The other participant (P7, VI) failed to finish a trial within the time limit (3 min.) due to the misrecognitions of the walk light detector. We observed that the participants in the VR group had proper orientation and camera movement speed while scanning the environment for 22.6% of the scanning time. On the other hand, the participants in the VI group did so only for 15.1% of the scanning time. These results suggest that the VR training method effectively allows BVI individuals to learn how to use a camera for a walk light detector.

Conclusions

In this study, we developed an interactive training app for BVI people where they practice interactions with a camera for a camera-based navigation system. To design the app, we interviewed BVI individuals to understand their challenges in using a camera for navigation. The interviews revealed that BVI people are interested in detecting landmarks and objects such as walk lights for navigation. They have challenges in image framing and taking clear photos with a camera. With these findings, we implemented an interactive training app with VR where BVI individuals can practice using a walk light detector with virtual walk lights. Through a pilot study, we found that interactive training is usable for BVI people, and we also encountered some user study design issues that we resolved for a user study. The intermediate results from the user study revealed that interactive training potentially enables BVI individuals to manipulate a camera better while they scan their environments to find a walk light.

Acknowledgments

JH was supported by Smith-Kettlewell's C.V. Starr Postdoctoral Fellowship. JMC was supported by NIDILRR grant number 90 REGE0018.

Works Cited

Afif, Mouna, et al. “An evaluation of retinanet on indoor object detection for blind and visually impaired persons assistance navigation.” Neural Processing Letters 51.3 (2020): 2265-2279.

Bigham, Jeffrey P., Jayant, Chandrika, Ji, Hanjie, Little, Greg, Miller, Andrew, Miller, Robert C., Miller, Robin, Tatarowicz, Aubrey, White, Brandyn, White, Samual, et al. 2010. *Vizwiz: nearly real-time answers to visual questions.* Proc. 23nd annual ACM symposium on User interface software and technology. 333–342.

Braun, Virginia and Clarke, Victoria. 2006. Using thematic analysis in psychology. Qualitative research in psychology 3, 2 (2006), 77–101.

Easwara Moorthy, Aarthi and Vu, Kim-Phuong L. 2015. Privacy concerns for use of voice activated personal assistant in the public space. International Journal of Human-Computer Interaction 31, 4 (2015), 307–335.

Jayant, Chandrika, Ji, Hanjie, White, Samuel, and Bigham, Jeffrey P. 2011. *Supporting blind photography*. Proc. 13th international ACM SIGACCESS conference on Computers and accessibility. 203–210.

Kreimeier, Julian, Karg, Pascal, and Götzelmann, Timo. “BlindWalkVR: formative insights into blind and visually impaired people's VR locomotion using commercially available approaches.” Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments. 2020.

Kunz, Andreas, et al. “Virtual navigation environment for blind and low vision people.” International Conference on Computers Helping People with Special Needs. Springer, Cham, 2018.

Lee, Kyungjun, Hong, Jonggi, Pimento, Simone, Jarjue, Ebrima, and Kacorri, Hernisa. 2019. *Revisiting blind photography in the context of teachable object recognizers.* 21st International ACM SIGACCESS Conference on Computers and Accessibility. 83–95.

Manduchi, Roberto and Coughlan, James M., 2014. *The last meter: blind visual guidance to a target.* Proc. SIGCHI Conference on Human Factors in Computing Systems. 3113–3122.

Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.

Serrão, Miguel, et al. "Indoor localization and navigation for blind persons using visual landmarks and a GIS." Procedia Computer Science 14 (2012): 65-73.

Tapu, Ruxandra, Mocanu, Bogdan, and Zaharia, Titus. "A computer vision system that ensure the autonomous navigation of blind people." 2013 E-Health and Bioengineering Conference (EHB). IEEE, 2013.

Theodorou, Lida, Massiceti, Daniela, Zintgraf, Luisa, Stumpf, Simone, Morrison, Cecily, Cutrell, Edward, Tobias Harris, Matthew, and Hofmann, Katja. 2021. Disability-First Dataset Creation: Lessons from Constructing a Dataset for Teachable Object Recognition with Blind and Low Vision Data Collectors. 23rd International ACM SIGACCESS Conference on Computers and Accessibility (Virtual Event, USA) (ASSETS '21). Association for Computing Machinery, New York, NY, USA, Article 27, 12 pages.

Vázquez, Marynel and Steinfeld, Aaron. 2012. *Helping visually impaired users properly aim a camera.* Proc. 14th international ACM SIGACCESS conference on Computers and accessibility. 95–102.

An Accessible BLE Beacon-based Indoor Wayfinding System

Ajay Abraham, Vinod Namboodiri
School of Computing, Wichita State University
axabraham1@shockers.wichita.edu, vinod.namboodiri@wichita.edu

Abstract

Navigating large indoor environments can be challenging, especially for those with disabilities. To address this issue, a Bluetooth Low Energy (BLE) beacon-based indoor wayfinding system has been developed, considering the unique needs of individuals with and without disabilities. The system was designed using the same methodology as the GuideBeacon indoor wayfinding system and features mobile applications for both Android and iOS devices. Based on usability testing conducted with 16 subjects across 4 user categories (including individuals with visual impairments, mobility impairments, hard of hearing, and those without disabilities), the indoor wayfinding system was found to be effective for individuals with diverse abilities.

Keywords

BLE beacon, localization, indoor navigation, people with disabilities.

Introduction

People with disabilities face several mobility challenges that prevent them from being engaged citizens. These include navigating independently in and to buildings that provide important services (educational services, medical care, etc.), and exploring unfamiliar environments safely and independently. Current navigation services fall short of addressing the wayfinding and navigation needs and preferences of people with disabilities.

For outdoor environments, a satellite-based system such as GPS provides a quick and easy means of wayfinding that is mostly accurate. However, in most indoor environments, GPS is neither reliable nor accurate. Reading indoor maps of buildings and following static signage is still the most reliable option but this indoor wayfinding strategy fails to consider the unique needs of people with disabilities. Even individuals without disabilities can struggle to navigate large, unfamiliar indoor environments. This could be due to the lack of well-marked signs and maps, not being familiar with the conventions or language used on the signage, or just the fact that the layout of the space is disorienting. Thus, there is a need for a low-cost, easy-to-use, and reliable wayfinding system that can aid both individuals with and without disabilities.

The early systems for enabling auxiliary location-based services (ALBSs) for independent wayfinding in GPS-limited areas were based on radio-frequency identification (RFID) tags (Wang and Katabi; Mooi et al.; Willis and Helal; Chang et al.; Kulyukin et al.) or infrared technology (Aitenbichler and Mühlhäuser). These were not very flexible for changing embedded information and required special hardware to be carried around by users. Such constraints had hampered the widespread use and adoption of ALBSs. Some recent efforts for wayfinding for the blind and visually impaired (BVI) using Bluetooth Low Energy (BLE) beacons that interact wirelessly with smartphones include StaNavi (Kim et al.), GuideBeacon

(Cheraghi et al.), ASSIST (Nair et al.), PERCEPT (Ganz et al.), and NavCog (Ahmetovic et al.). All report significant improvement in the ability of BVI individuals to navigate indoor spaces independently. Such beacon-based navigation systems may be a viable solution for indoor wayfinding for the BVI if the underlying challenges to their deployment can be overcome.

The biggest limitation of existing work is that it has targeted solving challenges for just one type of disability (such as vision loss), resulting in a very small potential user base. This work aims to be inclusive in addressing the unmet auxiliary wayfinding needs of people with and without disabilities using a common technology infrastructure.

Discussion

This work presents a wayfinding system that builds upon the methods developed for GuideBeacon, a BLE beacon-based indoor wayfinding system (Cheraghi et al.). Figure 1 shows the components of the BLE beacon-based wayfinding system. Every point of interest in an indoor setting that the user may wish to navigate to is associated with a BLE beacon. Each beacon is linked to at most 4 points of interest that lie close to each other. These beacons and the paths that connect them can be depicted in a connected graph *G(V, E)*, where *V* is the set of vertices representing the beacons and E representing the paths that exist between the beacons. Each path that a user can take between any two beacons *u* and *v* has a certain weight *w*, which could be a function of the distance between the beacons and/or the ease at which one can move between these beacons. Figure 2 shows a connected graph where the beacons serve as vertices and the paths between beacons serve as edges (the walking paths). The round markers represent the beacon locations, and the square markers represent the points of interest. Each of the edges has a weight (in this case, the actual distance between the markers) that determines if a path needs to be part of the end-to-end route computed during navigation. A shortest path algorithm (such as Dijkstra's algorithm) helps find the shortest path between the source node (the beacon that the user is close to) and the destination node (the point of interest that the user intends to navigate to). All this graphical information that contributes to wayfinding is stored in a database and later retrieved using a mobile application.

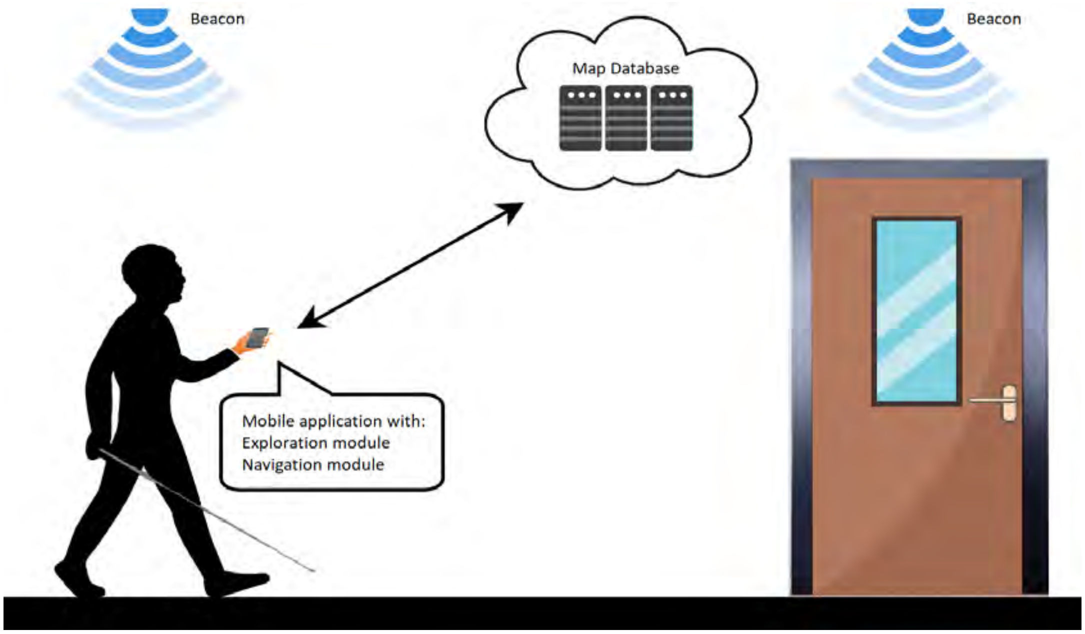


Fig. 1. BLE beacon-based wayfinding system.

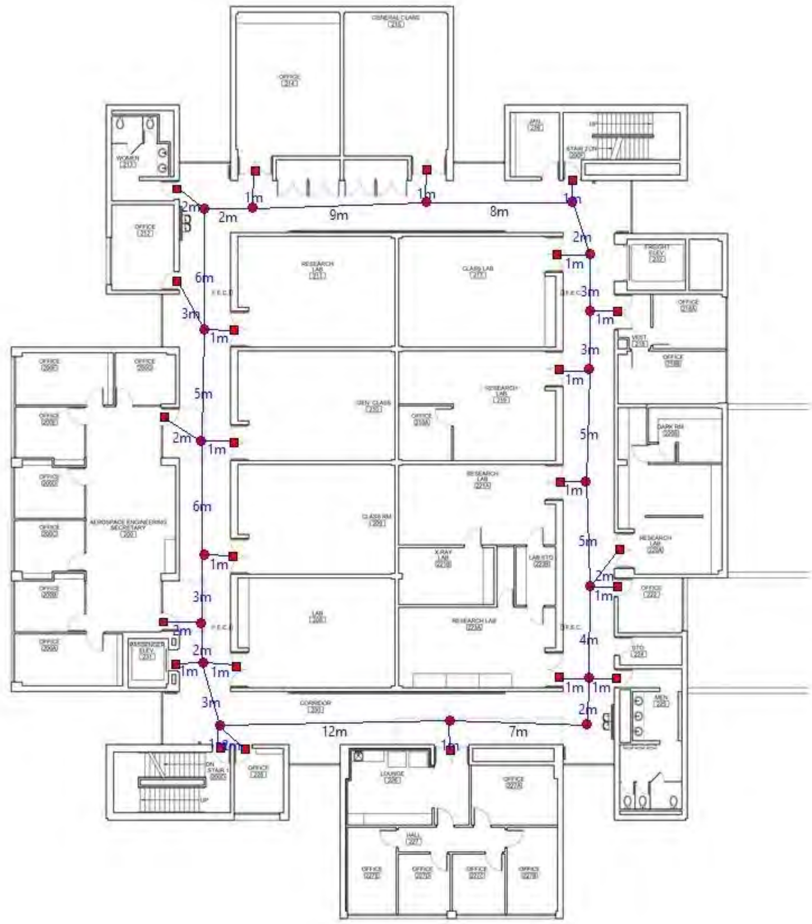


Fig. 2. Connected graph.

To estimate the position of a user in an indoor space, the wayfinding system utilizes a beacon-based proximity detection mechanism. It relies on received signal strength indicator (RSSI) readings to determine if a user is near a beacon; the closer the user is to a beacon, the higher the RSSI will be.

The mobile application developed for indoor wayfinding features two modules that can help with both localization and navigation. The exploration module activates when the user is near a BLE beacon that is part of the indoor beacon network. This module provides information about the nearby points of interest associated with a beacon. As the user moves through the indoor space and encounters new beacons, they will be able to discover new points of interest.

This is useful for users who simply wish to learn about their new surroundings without a specific destination in mind. While standing close to a beacon, users can perform a voice search or type in a search box to look up their desired destination in a database of indoor points of interest. Once the user chooses their destination, the navigation module activates. This module computes the end-to-end route and generates guidance instructions that help users navigate to their desired destination. If a user deviates from the computed path, a rerouting mechanism kicks in and computes a new route from the user's current location to the intended destination and guides the user along this new route.

Beacon Deployment

The three-story college campus building that houses the authors' research laboratory was chosen as the beacon deployment site. Its 2-meter wide, double-loaded corridors served as a testing ground for the BLE beacon-based indoor wayfinding system presented in this paper. Due to its low cost and long battery life, the Gimbal Series 21 beacon was the beacon of choice for this work. The Gimbal Series 21 beacon uses four standard AA alkaline batteries and has a battery life of around 18 months.

Beacons were installed on all three floors of the campus building. Beacon placement locations included entrances to rooms, corner spaces of angled corridors, and entryways to vertical circulation spaces such as elevators and stairwells. Beacons placed near elevators and stairwell doors served as floor transition beacons that allowed for navigation between floors.

Mobile Applications

Two mobile applications were developed for operating the wayfinding system presented in this paper: one for Android and another for iOS devices. Both these applications can seamlessly switch between the exploration module (which allows users to learn about their immediate surroundings) and the navigation module (which provides navigational guidance to users as they move toward their destination).

The mobile applications were equipped with a feature set that is largely the same. The route preview feature gives users a preview of the entire route (which includes information about the turns and the distances to these turns) from the user's current position to their destination. Both applications were equipped with built-in accessibility features of their respective platforms. Users can customize settings, which include changing the audio narration speed, turning the route preview on or off, changing the unit for distance to the next point on the route (meters, feet, or steps), and changing how relative directions are given (left-right-forward or clock position). Within the application, users can look up a destination in the database of indoor points of interest either by typing text in a search box or by using voice search (user input captured through voice is converted to text using a text-to-speech application). For both these types of input, users can search through partial or other known information associated with a destination. For example, the room number associated with a location could be used in place of the location name to search the database.

The differences between the iOS and Android applications lie in the presentation of the system responses and navigational instructions. In addition to voice directions, users are provided turn-by-turn directions in a list format on the screen in the Android application (Figure 3a). This list always displays the upcoming turn instruction as the first item and shortens as the user approaches their destination. In the iOS application, the user has to rely on voice directions for navigation, but the user's position will be denoted as a dot on the floor plan that updates frequently as the user moves towards their destination (Figure 3b). This feature helps the user track their location in the indoor space at any given time. Users also have the option to double-tap anywhere on the screen to repeat the last spoken turn-by-turn direction (while in navigation mode) or to replay the audio description of their surroundings (while in exploration mode).

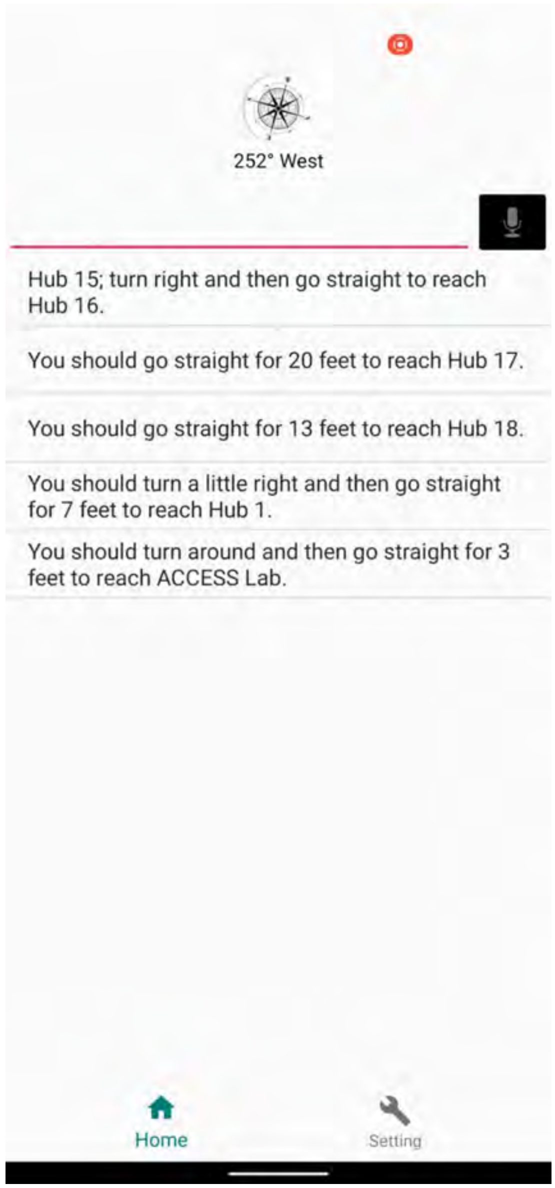


Fig. 3a. Android application. Guidance provided by the mobile applications when the navigation module is active.

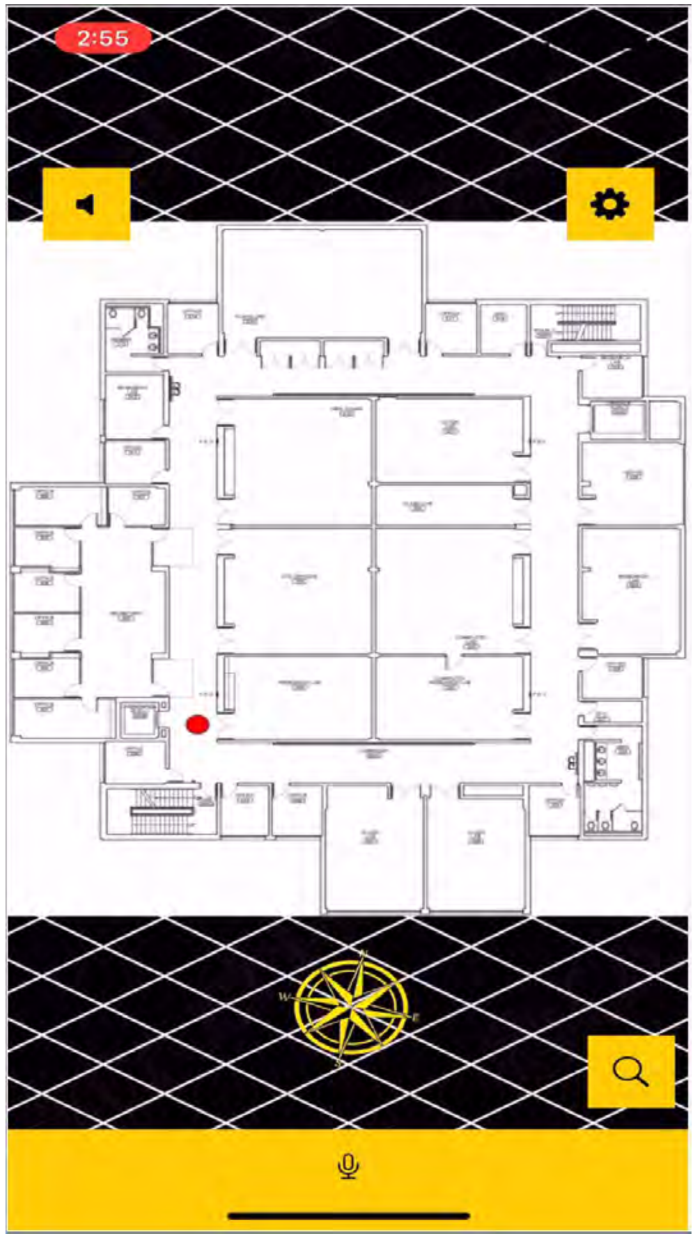


Fig. 3b. iOS application. Guidance provided by the mobile applications when the navigation module is active.

Usability Testing

A total of 16 subjects who were unfamiliar with the building layout of the test site were recruited for testing the wayfinding system: 4 with visual impairments (cane users), 4 with mobility impairments (electric wheelchair users), 4 who are hard of hearing, and 4 with no disabilities. From each of the four user categories, two were asked to navigate without any assistance aside from their mobility aid (the control group), and the other two were asked to navigate using both the wayfinding app and their mobility aid (the experimental group). Each test subject was first positioned near the accessible entrance on the ground floor and then asked to make their way to the authors' research lab on the third floor. This test route is illustrated in Figure 4. All test subjects were asked to take the elevator to get to the third floor. To evaluate the effectiveness of the navigation module of the wayfinding app, the navigation time and navigation distance for each test subject were recorded. To evaluate the performance of the exploration module of the wayfinding app, the test subjects were asked to walk around and explore the third floor of the building.

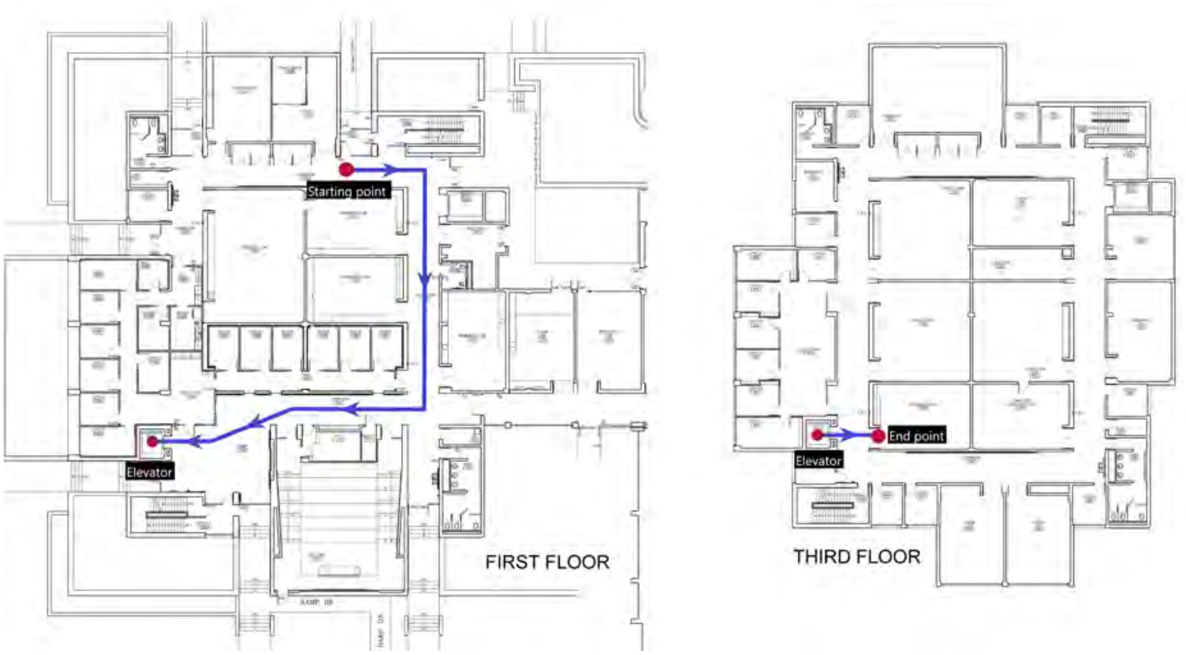


Fig. 4. Test route.

The quantitative results obtained from this testing are shown in Figures 5 and 6. During testing, the time spent waiting for and riding the elevator was not considered when calculating the navigation time. Within each user category, the test subjects who used the wayfinding app took less time to reach their destination and traversed less distance overall when compared to the test subjects who navigated the test route without the app. For all user categories excluding the BVI, the difference in results between the control group and the experimental group does not seem significant. This is because the test route used only had one main intersection where the subjects had to decide what path to take. None of the test subjects who used the wayfinding app took a wrong turn. Among the test subjects who did not use the app, the ones who took a wrong turn had to backtrack which resulted in an increase in their navigation distance and navigation time. Test subjects with visual impairments faced different challenges compared to their sighted counterparts. While the latter could visually verify if they had made a wrong turn and check for the elevator's presence, those with visual impairments had to rely on manual search and identification of each point of interest along their path (such as reading braille door signs) to ensure they did not miss the elevator. This significantly increased the navigation time and distance for individuals with visual impairments.

Qualitative data was collected using a combination of observational usability testing and post-test interviews. When exploring unknown indoor spaces, the test subjects with visual impairments who used the app were satisfied with the exploration module since they were able to discover all points of interest as they moved along the corridor path. The ones with visual impairments who did not use the app took considerable time and effort in finding and identifying the points of interest along the path since they had to resort to touching the walls and interpreting the braille signage. Those with no visual impairments did not find listening to information about their immediate surroundings (the names of the points of interest nearby) to be particularly useful since each of the points of interest had a sign near its entrance that clearly showed the room name. However, having a map of the indoor space in the form of a digital floor plan on the wayfinding app gave them a bird's eye view of all points of interest and helped them learn about the indoor environment more effectively. Additionally, the dot on the floor plan indicating their current position in relation to other points of interest was beneficial. All test subjects found the double-tap feature that allowed them to listen to the same information more than once to be helpful.

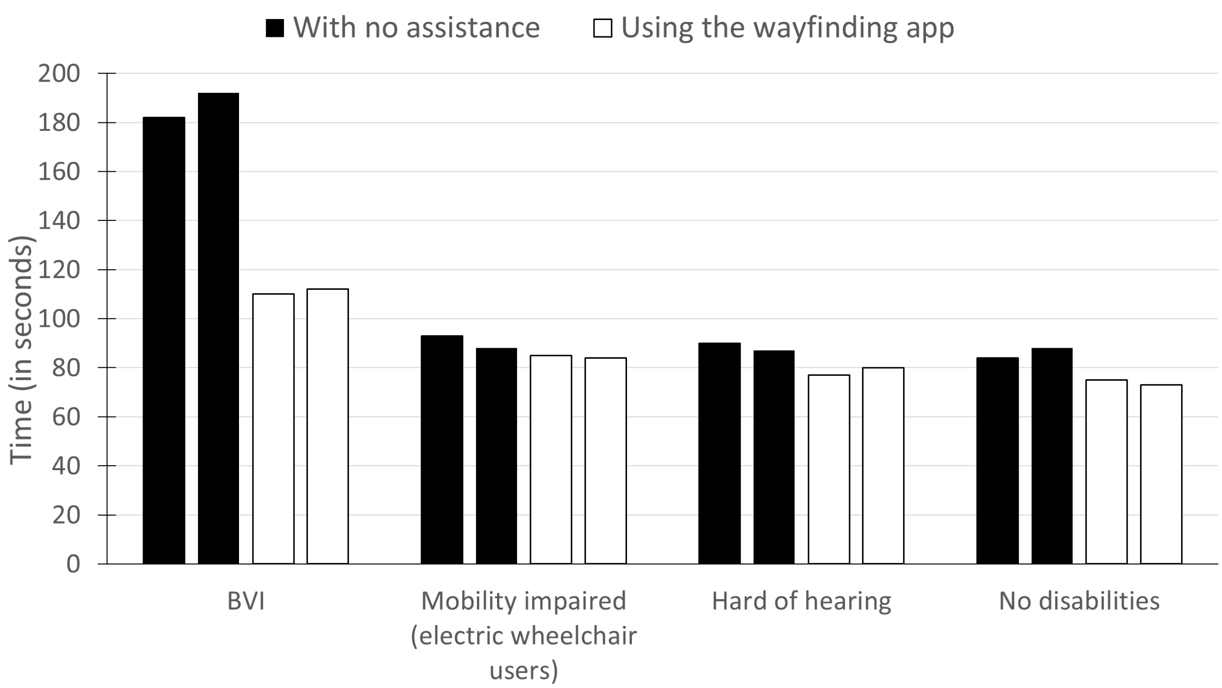


Fig. 5. Time taken by test subjects to navigate the test route.

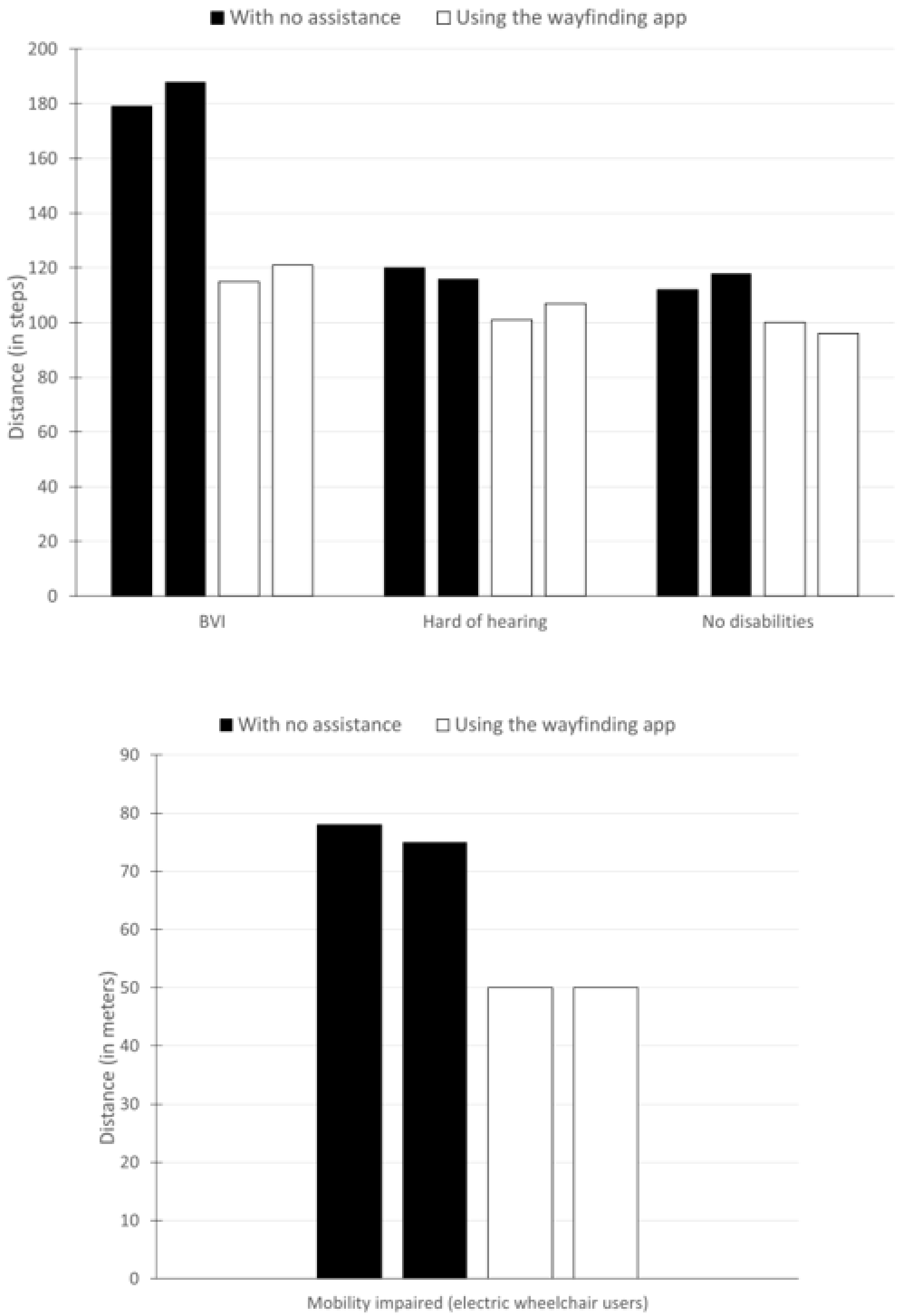


Fig. 6. Distance traversed by test subjects from the starting point to the endpoint of the test route.

When using the mobile application, almost all BVI users wanted information relayed to them at a very fast rate and would therefore set the application's narration speed to a high value. This confirmed that most BVI individuals can easily understand speech at a fast rate.

The wayfinding app successfully met the unique navigation requirements of all four user categories tested. While the subjects who are hard of hearing wanted turn-by-turn navigation instructions to be displayed as text on the screen, others wanted voice navigation. When looking up possible indoor destinations to navigate to, most BVI subjects preferred using voice search instead of manually entering text in the search box. Since they had a smartphone in one hand and a mobility aid such as a white cane in the other, they wanted hands-free control of the application. Most wheelchair users favored this hands-free operation as well. All subjects, irrespective of the user category they belonged to, found the navigation module of the mobile application to be the most useful.

Conclusions

The BLE beacon-based indoor wayfinding system presented in this work considers the unique information needs of people both with and without disabilities. The usability testing results suggest that people with diverse abilities can use the indoor wayfinding system effectively.

Works Cited

Ahmetovic, Dragan, et al. "NavCog: A Navigational Cognitive Assistant for the Blind." *International Conference on Human Computer Interaction with Mobile Devices and Services*. ACM, 2016.

Aitenbichler, Erwin, and Max Mühlhäuser. "An IR Local Positioning System for Smart Items and Devices." *Proceedings of the 23rd International Conference on Distributed Computing Systems*. ICDCSW '03, IEEE Computer Society, 2003, pp. 334–.

Chang, Yao-Jen, et al. "Autonomous indoor wayfinding for individuals with cognitive impairments." *Journal of NeuroEngineering and Rehabilitation*, vol. 7, no. 1, 2010, pp. 1–13.

Cheraghi, Seyed Ali, et al. "GuideBeacon: Beacon-Based Indoor Wayfinding for the Blind, Visually Impaired, and Disoriented." *IEEE International Conference on Pervasive Computing and Communications*. Mar. 2017.

Ganz, Aura, et al. "PERCEPT navigation for visually impaired in large transportation hubs." *J. Technol. Persons Disabilities*, Mar. 2018, pp. 336–53.

Kim, Jee-Eun, et al. "Navigating Visually Impaired Travelers in a Large Train Station Using Smartphone and Bluetooth Low Energy." *Proceedings of the 31st Annual ACM Symposium on Applied Computing*. SAC '16, 2016, pp. 604–11.

Kulyukin, Vladimir, et al. "Robot-assisted wayfinding for the visually impaired in structured indoor environments." *Autonomous Robots*, vol. 21, no. 1, 2006, pp. 29–41.

Mooi, Wong Siew, et al. "Efficient RFID tag placement framework for in building navigation system for the blind." *Information and Telecommunication Technologies (APSITT), 2010 8th Asia-Pacific Symposium on*. June 2010, pp. 1–6.

Nair, Vishnu, et al. “ASSIST: Personalized indoor navigation via multimodal sensors and high-level semantic information.” *The European Conference on Computer Vision (ECCV) Workshops*. Sept. 2018.

Wang, Jue, and Dina Katabi. “Dude, Where’s My Card?: RFID Positioning That Works with Multipath and Non-line of Sight.” *Proceedings of the ACM SIGCOMM 2013 Conference on SIGCOMM*. SIGCOMM ’13, 2013, pp. 51–62.

Willis, Scooter, and Sumi Helal. “RFID Information Grid for Blind Navigation and Wayfinding.” *Proceedings of the Ninth IEEE International Symposium on Wearable Computers*. ISWC ’05, 2005, pp. 34–37.

Live Captions in Virtual Reality (VR)

Pranav Pidathala, Dawson Franz, James Waller, Raja Kushalnagar, Christian Vogler
Gallaudet University
pranav.pidathala@gallaudet.edu, dawson.franz@gallaudet.edu, james.waller@gallaudet.edu, raja.kushalnagar@gallaudet.edu, christian.vogler@gallaudet.edu

Abstract

Few VR applications and games implement captioning of speech and audio cues, which either inhibits or prevents access of their application by deaf or hard of hearing (DHH) users, new language learners, and other caption users. Additionally, little to no guidelines exist on how to implement live captioning on VR headsets and how it may differ from traditional television captioning. To help fill the void of information behind user preferences of different VR captioning styles, we conducted a study with eight DHH participants to test three caption movement behaviors (head-locked, lag, and appear-locked) while watching live-captioned, single-speaker presentations in VR. Participants answered a series of Likert scale and open-ended questions about their experience. Participants' preferences were split, but most participants reported feeling comfortable with using live captions in VR and enjoyed the experience. When participants ranked the caption behaviors, there was almost an equal divide between the three types tested. IPQ results indicated each behavior had similar immersion ratings, however participants found head-locked and lag captions more user-friendly than appear-locked captions. We suggest that participants may vary in caption preference depending on how they use captions, and that providing opportunities for caption customization is best.

Keywords

Virtual reality caption access; deaf and hard of hearing; automatic speech recognition accessibility; virtual reality subtitle usability.

Introduction

Virtual Reality (VR) as a mainstream industry is an extremely recent development, with today's technology only becoming widely available around six years ago with the public release of VR headsets like the Oculus Rift, HTC Vive, and others. In virtual reality, users have a 360º view of their surroundings: they can freely view everything in the virtual world by moving either their head or a controller. However, with novel technology comes novel accessibility challenges.

VR technology that uses spoken audio must provide that information in an alternative format to be accessible to Deaf and Hard of hearing (DHH) users – such as through captions. Captions typically appear at the bottom of a screen during the duration of the spoken audio, with standard guidelines such as a black background box, sans-serif white font, and a maximum of 3 lines with 30-40 characters per line.

However, virtual reality adds several new dimensions to consider. For example, when the person moves, should the captions move with them? Or do they stay fixed in place, anchored to some landmark in the virtual environment? A crucial element of virtual reality is its sense of immersion, so it is import ant to design captions that are well integrated and do not break the immersive experience. Previous work has identified several different approaches to caption positioning in VR, which can be divided into two approaches – 1) 'head-locked,' and 'world-locked' (Brown et al., 2017). In the first approach, characterized as 'head-locked' captions, captions are consistently locked to a fixed position in the user's view. If the user shifts their head, the captions will follow. The captions stay in the same position within the user's view; they are always visible. One version of head-locked captions has a lag: the captions follow the head movements with a slight delay (allowing for smoother movement).

In contrast, in the ‘world-locked’ approach, the captions are not locked to a fixed position in the person’s *view*, but rather are ‘locked to the environment’. Just like other virtual objects, these captions stay fixed in the environment as the person moves around. If the user turns away from the captions, the captions disappear from view; they do not follow the user’s head movements. Some are ‘speaker/source-locked’ – meaning they appear near the source of the audio (usually below the speaker), or near an object associated with the caption content. Others, like ‘appear’ captions, use a hybrid approach between head-locked and world-locked captions. Appear-locked captions are locked to the user’s initial head position, so that when the captions appear, they appear where the user is looking. But, unlike head-locked captions, they stay in place in the world if the user looks away.

Recent research has compared different captioning styles with various users. Rothe et al. (2018) compared hearing people’s preference for speaker-locked or head-locked captions in a VR setting with foreign language audio, and found no significant difference. Agulló (2019) tested head-locked captions without lag and ‘fixed’ (world-locked) captions with hearing and DHH participants. In the world-locked caption condition there were three different caption set ups spaced evenly around the user 120 degrees apart, such that wherever the user was looking there was likely at least one set of captions visible. They found that 82.5% of participants (DHH and hearing) strongly preferred the head-locked captions over the multiple world-locked captions. A follow-up study confirmed these findings, and found that participants rated the head-locked captions as more immersive (Agulló et al., 2020).

Research so far primarily centers on pre-recorded, pre-captioned 360^{o} video content. These approaches could be expanded upon to include other types of interactive VR media as well, such as in applications with live user-to-user interaction, but we found little to no mention of live VR captioning in past research. Social media, education, and virtual meeting platforms that support VR and feature active user interaction should implement live audio captioning as an accessibility option to keep their platform accessible for DHH users, yet few of the platforms we tested had this as an option. In our research on VR meeting platforms, we found that only a few supported live captioning, and those that did support it seemed to provide captioning as a paid service with little to no customizability options. We found one web-based immersive meeting platform in particular, Mozilla Hubs, which does not currently support captioning but is open source and can thus be modified by anyone. Based on this information, we decided to implement some of the caption types listed above within this platform to get a better idea of how practical they are to use and see how they would impact a DHH user's experience within VR.

Our study tested three different live VR caption types with DHH users: 1) head-locked captions without lag, 2) head-locked captions with lag, and 3) appear-locked captions in which the captions always appear in the user’s visual field but stay locked to the world thereafter. We exclusively recruited DHH users, who may have different preferences for VR captions than other users: many DHH users are frequent caption users, and do not necessarily rely on audio information to know if someone is speaking at that moment.

Discussion

For the present study, we tested three types of captioning behaviors in VR with Deaf and hard of hearing participants in the U.S.

Implementation

We coded prototypes of the three different captioning behaviors in Mozilla Hubs, an open-source, web-based meeting platform that supports VR headsets through the WebVR application interface. We implemented three different caption conditions, as follows:

'Head-locked captions' (no lag)

- Always visible.
- Moves along x-axis only. Locked on the y-axis to avoid blocking content: it cannot move up or down.

'Lag captions'

- Usually always visible, except when the user moves quickly.
- Do not move up to down. When moving on the x-axis, the captions usually take approximately one second to 'catch up' to the user's movement.

'Appear-locked captions'

- The caption box is shown when the user is speaking. Once the speaker starts a new sentence, the captions go to where the user is looking.
- If the user is looking in the same spot continuously, the captions go under each other.

For consistency, all aspects of the captioning types other than their movement behavior are kept the same and use the same textbox object. We emulate existing TV closed captioning guidelines for our caption textbox: large white sans-serif text on a non-transparent black box with up to three lines of text (max. 128 characters). At all times, the caption box is facing the user.

We used Mozilla Hubs as our meeting platform, and Javascript as our primary coding language. The website platform was hosted on AWS. With our custom client and cloud server, we add onto Hubs' existing codebase using the A-frame framework for 3D Objects to support captioning. We populated the world, or scene, with various 3D background elements, such as stairs, elevated platforms, and images to give the user an environment to visually inspect and explore as they traveled through. Four separate world templates, all with the same layout and 3D objects but with different images, were created within Spoke, Mozilla Hubs' default scene editor, to go along with a corresponding educational script and presentation: the first was about the history of the Meta Quest VR device, the second was about the history of captioning devices, the third was about the Gallaudet Eleven's contributions to motion sickness research, and the fourth was a practice scene with placeholder images. A fourth non-moving caption type (permanently world-locked, located next to where the user enters the scene) was used only in our practice scene for acclimating the user to the VR environment and was not formally tested.

Captions were provided by Microsoft Azure's speech-to-text service for the live voice transcription through a W3C Speech Recognition polyfill so it works on all browsers. In our implementation on the Meta Quest 2 VR headset, Mozilla Hubs runs on the device through the Quest's default browser, where the speech recognition polyfill sends the audio directly from the microphone of the headset to Azure Cognitive Services. Captions are then sent back to the browser and formatted for readability. The goal is to mimic the use of VR for a live educational presentation or virtual tour.

Methodology

Before the experiment, the participants fill out a questionnaire asking about demographic information and experience with VR and captions. At the start of the experiment, the experimenter explains the instructions and controls. The participant puts on the VR headset, calibrates the equipment, and familiarizes themselves with a practice environment. In our study, participants could move around the scene using either Hubs' teleport mechanic (by holding down the left or right trigger button on the oculus controller and then releasing), or by using the left-joystick to "walk" around. They were informed about both movement types, with the note that joystick movement could cause motion sickness. Sometimes the captions stopped displaying text due to a bug, so participants were also informed how to mute and unmute the microphone on the Quest to fix the problem.

During the experiment, the participant experienced a ‘live educational presentation’, which lasted approximately five minutes. The experimenter gave a scripted presentation about different topics which are displayed in the VR environment, such that the participant could wander around the displays as the experimenter discussed each in turn. Automatically generated captions would appear in the participant’s VR environment while the presenter was talking.

We did this three times, once for each caption condition. Each participant saw each scene once, and caption ordering was counterbalanced with the scene ordering to remove bias. After each caption condition, the participant took off the VR headset and answered a series of 7-point Likert scale questions regarding their immersion in the VR world (using the IPQ questionnaire, see Regenbrecht and Schubert, 2002), as well as how easy and user-friendly the captions were to use. Participants also had the opportunity to provide open-ended feedback, and at the end of the tests they were able to rank the three caption behaviors from favorite to least favorite.

Eight DHH participants participated in our study, with ages ranging from 19 years old to 63 years old. Four were women, three were men, and one was non-binary. Six of the participants identified as deaf, and two as hard of hearing. Most of the participants used ASL as their main form of communication. Some people preferred ASL along with written English.

All participants used captions regularly (every day or most days) – most commonly for watching TV or streaming videos, but also for education, webinars, videoconferencing, and phone calls. Six participants use captions as their “primary method to understand information” and two use them as their secondary method and “to catch a few words I missed.” All participants had little to no previous experience with VR.

When asked if they used hearing devices, six participants did not currently use any, one uses a cochlear implant, and one uses a hearing aid. Five of the participants had glasses and wore them during the study, and three did not. None of the participants identified as DeafBlind.

Results

Participants were split in their caption preferences: three preferred head-locked captions, two preferred lag captions, and two preferred the appear-locked captions. In their comments, several participants said they liked that the head-locked and lag captions were consistently in the same place for reading the captions (both locked to head position) – though for others the captions were too low or too close, which was a source of frustration. Conversely, those who favored appear-locked captions liked that they moved relative to the head and did not stay in one position. Both participants who preferred the appear-locked captions use hearing devices and used captions as their secondary method of understanding information. Some participants noted frustrations about the caption types. For head-locked and lag captions, many participants mentioned the captions were too low and some participants wished the captions followed up and down with their head as well as side to side.

There were some complaints about the position of the appear-locked captions – too low, or too close. Also, one participant notes the captions sometimes "obscured the picture [they were] looking at". Another participant experienced "caption drift" with appear-locked captions, where the captions kept moving downward when they turned their head to read the text.

Positive comments were mostly directed at head-locked and lag captions. Many participants found them clearer than the appear-locked captions – the captions were always readable without having to look back at a specific place. One participant also commented the consistency in head-locked captions allows them to also view the environment better: “The consistency in place and movement made it easier to read and focus on the environment.”.

Overall, any issues that arose with the captions did not seem to turn participants off from the technology as a whole – post-study, all but one participant reported feeling extremely (n=4) or moderately (n=3) comfortable in VR, and several commented that they enjoyed the experience and felt immersed.

When asked to rate the user friendliness of the caption type (1 to 7 scale: 1 = Worst imaginable, 4 = OK, 7 = Best), appear-locked captions scored the lowest (M = 3.9) while the scores were similar for head-lock captions (M = 5.2) and lag captions (M = 5.0). The three caption types all scored between 4 and 5 on immersion (scores calculated from the IPQ, on a scale from 1 to 7).

General Discussion

DHH users are not uniform in their preferences, and it seems a flexible, customizable approach is needed. Not only did participants favor different caption types, but some suggested alternative approaches that allowed even more control over the captions, such as being able to drag and pin the captions to any position in the environment: “I would like to be able to either ‘pin’ the captions in a specific spatial location in my vision field (for head-lock approach). For the appear-locked captions, it needs to be [interact-able] so I can move it next to the pictures or the other [visual information] of interest.”

Caption preferences appear to vary based on how people use captions and the VR technology. For example, people who rely on captions as a secondary source of information may like to be able to ‘look away’ from captions, but others who rely primarily on captions prefer head-locked captions that are always visible.

Caption preferences may also depend on the way that users move around their virtual environment. For example, we observed that one participant complained about appear-locked captions being ‘too close’ – they also were one of the few used the teleport function to move around. It is possible that the teleport function offers less fine control of positioning than the joystick, so it is more frustrating to have to re-position relative to the appear-locked captions for those using teleport movement. Future research should account for the possibility that the ideal caption set-up depends on how users navigate the virtual environment – whether through joystick, teleportation, or physical movement (not tested in this study): i.e., walking around the physical environment to navigate the virtual one.

Conclusions

Compared to TV closed captions, live VR captions may require additional customization options and/or should exist as an interactable object, since the expanded view of VR may not allow for a one-size-fit-all approach. When developing for VR, it is a good idea to give users multiple accessibility options, along with the tools to customize them to what's best for them.

This study focused on the use of live captions in virtual reality rather than pre-recorded captions. While principles developed for live captions in VR may be largely applicable to pre-recorded captions, presentation of pre-recorded captions can also be tailored in ways that cannot be easily done with live captions – such sophisticated source-locked caption options that we did not test here. Additionally, pre-recorded captions could also include information such as emotion and tone of voice, or environmental sounds. How best to integrate this information into VR applications is still an open question.

Nonetheless, current immersive captioning work is promising, and shows that captions do not necessarily detract from the VR experience – many DHH participants are excited to participate in VR technology, and it is important that these technologies are accessible to all.

Acknowledgments

This work was supported by NSF #2150429 (REU AICT).

Works Cited

Agulló, Belén, and Anna Matamala. “Subtitling for the deaf and hard-of-hearing in immersive environments: results from a focus group.” *The Journal of Specialised Translation* 32 (2019): 217-235.

Agulló, Belén, and Anna Matamala. “Subtitles in virtual reality: guidelines for the integration of subtitles in 360º content.” Íkala, revista de lenguaje y cultura 25.3 (2020): 643-661.

Brown, Andy, et al. “Subtitles in 360-degree Video.” *Adjunct Publication of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video*. 2017.

Regenbrecht, Holger, and Thomas Schubert. “Real and illusory interactions enhance presence in virtual environments.” *Presence: Teleoperators & Virtual Environments* 11.4 (2002): 425-434.

Rothe, Sylvia, Tobias Höllerer, and Heinrich Hußmann. “CVR-analyzer: a tool for analyzing cinematic virtual reality viewing patterns.” *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia* (2018): 127-137.

ASL Consent in the Digital Informed Consent Process

Ben S. Kosa[1], Ai Minakawa[2], Patrick Boudreault[2], Christian Vogler[2], Poorna Kushalnagar[2], Raja Kushalnagar[2]
University of Washington[1]
Gallaudet University[2]
ben.kosa@gallaudet.edu, ai.takayama@gallaudet.edu, patrick.boudreault@gallaudet.edu, christian.vogler@gallaudet.edu, poorna.kushalnagar@gallaudet.edu, raja.kushalnagar@gallaudet.edu

Abstract

There is an estimated 500,000 people in the U.S. who are deaf and who use ASL and live in the U.S. Compared to the general population, deaf people are at greater risk of having chronic health problems and experience significant health disparities and inequities (Sanfacon, Leffers, Miller, Stabbe, DeWindt, Wagner, & Kushalnagar, 2020; Kushalnagar, Reesman, Holcomb, & Ryan, 2019; Kushalnagar & Miller, 2019). Much of the disparities are explained by the barriers in the environment, such as the unavailability of materials in ASL and lack of healthcare professionals who know how to provide deaf patient-centered care. Intersecting social determinants of health (e.g., intrinsic - low education; and extrinsic - barrier to healthcare services) create a mutually constituted vulnerability for health disparities when a person is deaf (Kushalnagar & Miller, 2019; Lesch, Brucher, Chapple, R., & Chapple, K., 2019; Smith & Chin, 2012). Moreover, the longstanding history of inequitable access to language and education, and a lack of printed information and materials, leave people who are deaf and use ASL unaware of opportunities to participate in cutting-edge research/clinical trials. An unintended consequence, therefore, is that PIs neglect to include people who are deaf and use ASL in their subject sample pools, and this marginalized population continues to be at disparity for health outcomes and also clinical research participation. One barrier is the unavailability of informed consent materials that are accessible in ASL.

Continue….

THE JOURNAL ON TECHNOLOGY AND PERSONS WITH DISABILITIES

ASL Consent in the Digital Informed Consent Process

Ben S. Kosa[1], Ai Minakawa[2], Patrick Boudreault[2], Christian Vogler[2], Poorna Kushalnagar[2], Raja Kushalnagar[2]
University of Washington[1]
Gallaudet University[2]
ben.kosa@gallaudet.edu, ai.takayama@gallaudet.edu, patrick.boudreault@gallaudet.edu, christian.vogler@gallaudet.edu, poorna.kushalnagar@gallaudet.edu, raja.kushalnagar@gallaudet.edu

Continued…

The current research study conducted by our team at the Center for Deaf Health Equity at Gallaudet University attempts to address the language barrier to the consent process through a careful reconsideration of its traditional English format and the development of an American Sign Language (ASL) informed consent app. This team successfully leveraged existing machine learning methods to develop a way to *navigate* and *signature* an informed consent process using ASL. We call this new method of navigation and signature "*ASL consent.*" In our findings, we found that deaf people who are primarily college educated were more likely to agree that the process for obtaining *ASL consent* through an accessible app is comparable to traditional English consent.

Keywords

Deaf and hard of hearing, American Sign Language, Consent, Machine Learning.

Introduction

There is an estimated 500,000 people in the U.S. who are deaf and who use ASL and live in the U.S. Compared to the general population, deaf people are at greater risk of having chronic health problems and experience significant health disparities and inequities (Sanfacon, Leffers, Miller, Stabbe, DeWindt, Wagner, & Kushalnagar, 2020; Kushalnagar, Reesman, Holcomb, & Ryan, 2019; Kushalnagar & Miller, 2019). Much of the disparities are explained by the barriers in the environment, such as the unavailability of materials in ASL and lack of healthcare professionals who know how to provide deaf patient-centered care. Several researchers have conducted focus group discussions with deaf people who use ASL and identified many barriers and facilitators involved in the informed consent process (Singleton, Jones, & Hanumantha, 2014; Anderson et al., 2020). They also articulate a stated need to train researchers and interpreters on the topic of cultural competence when working with deaf people in research. As a result, two informed consent and research participation training products resulted from this earlier research; one ASL video targets deaf people who wish to learn more about informed consent, the other video targets hearing researchers who have little experience or working knowledge of deaf people who use ASL.

Reconsidering Traditional Informed Consent Procedures

The informed consent process ensures that participants understand the information and risks necessary to decide whether they want to participate in the research. However, in almost all research, the informed consent process is done entirely in written or spoken English with a reading level of high school or above. Often, deaf people’s first natural language is ASL and they only learn written English as a second language later in life; this means that English literacy varies widely among deaf people who use ASL. This means most informed consent processes are shown with a reading level higher than most deaf AND hearing participants can understand, as research also suggests an eighth-grade average reading level amongst high-school graduates in the U.S. (Easterbrooks, 2012).

Informed consent forms are also widely inaccessible due to their high use of health-related jargon, as a large majority of deaf sign language users have low-health literacy due to limited language access during key developmental periods and “a lifetime of limited access to information that is often considered common knowledge among hearing persons” (Anderson et al, 2020; McEwen & Anton-Culver, 1988, Kushalnagar 2017). Most informed consent protocols, which often include written English informed consent forms with complex jargon and legal terminology, fail to fully inform a deaf research participant and call for adaptations above-and-beyond simple translation of informed consent materials into ASL (Kushalnagar, 2018, Kushalnagar 2020).

To address this, the Center for Deaf Health Equity team at Gallaudet University received funding from the National Institutes of Health to develop a user-centered and low-resource informed consent toolkit that allows developers and researchers to easily create informed consent applications that are accessible to deaf people who use ASL. In order to develop this toolkit, two things need to first be proven: the informed consent process is more accessible when its **(1)** *content* is shown in ASL and **(2)** when it’s *navigable* and *signaturable* in ASL.

As such, this paper aims to answer the following question: If the informed consent process in a research survey app instead allowed participants to navigate and consent in ASL, do deaf people who use ASLview this as equitable to the traditional informed consent process in-terms of **(1)** comprehension of what they are consenting to and **(2)** user friction in the consent process? This paper attempts to answer this question.

Fig. 1. A labeled English section in the ASL-Consent app with ASL Signature.

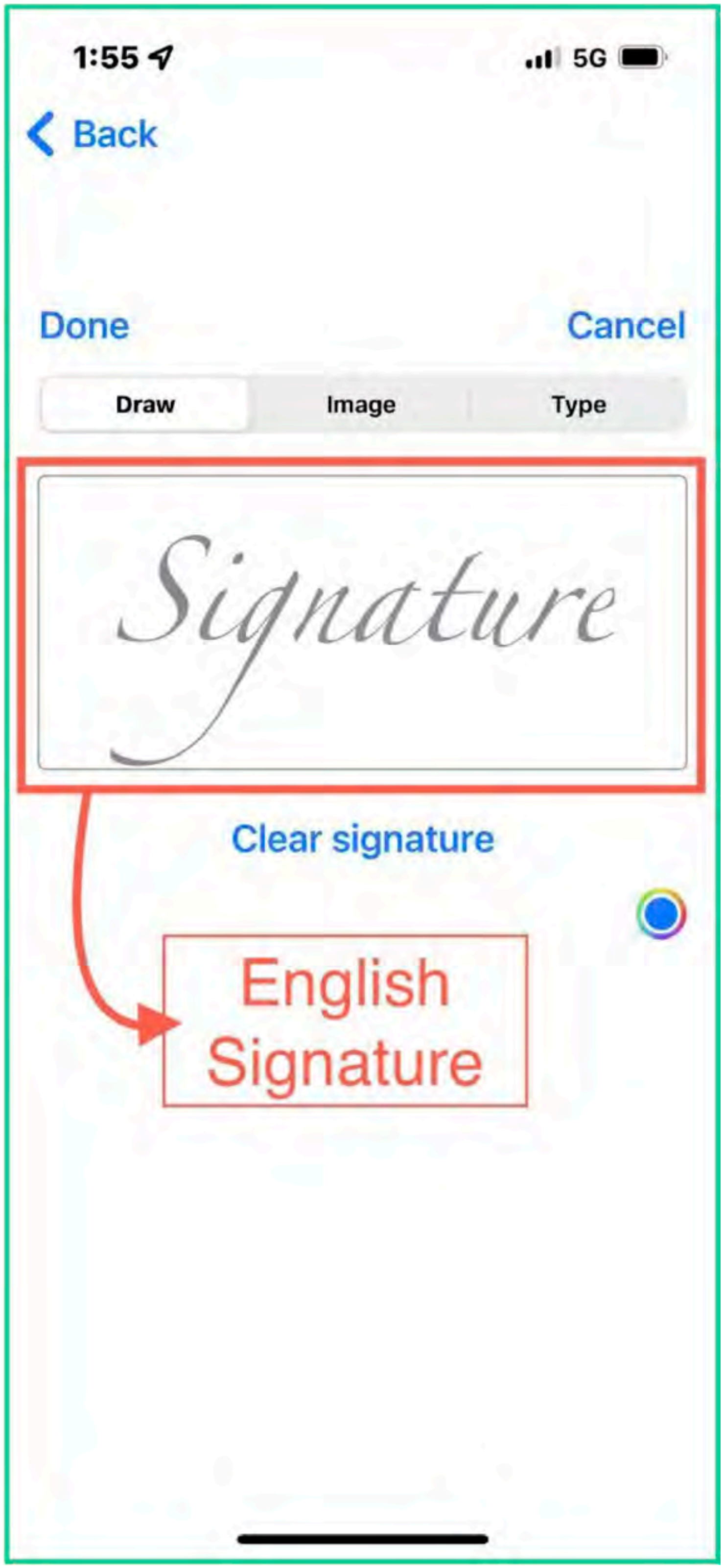


Fig. 2. A labeled signature section in the ASL-Consent app with English signature.

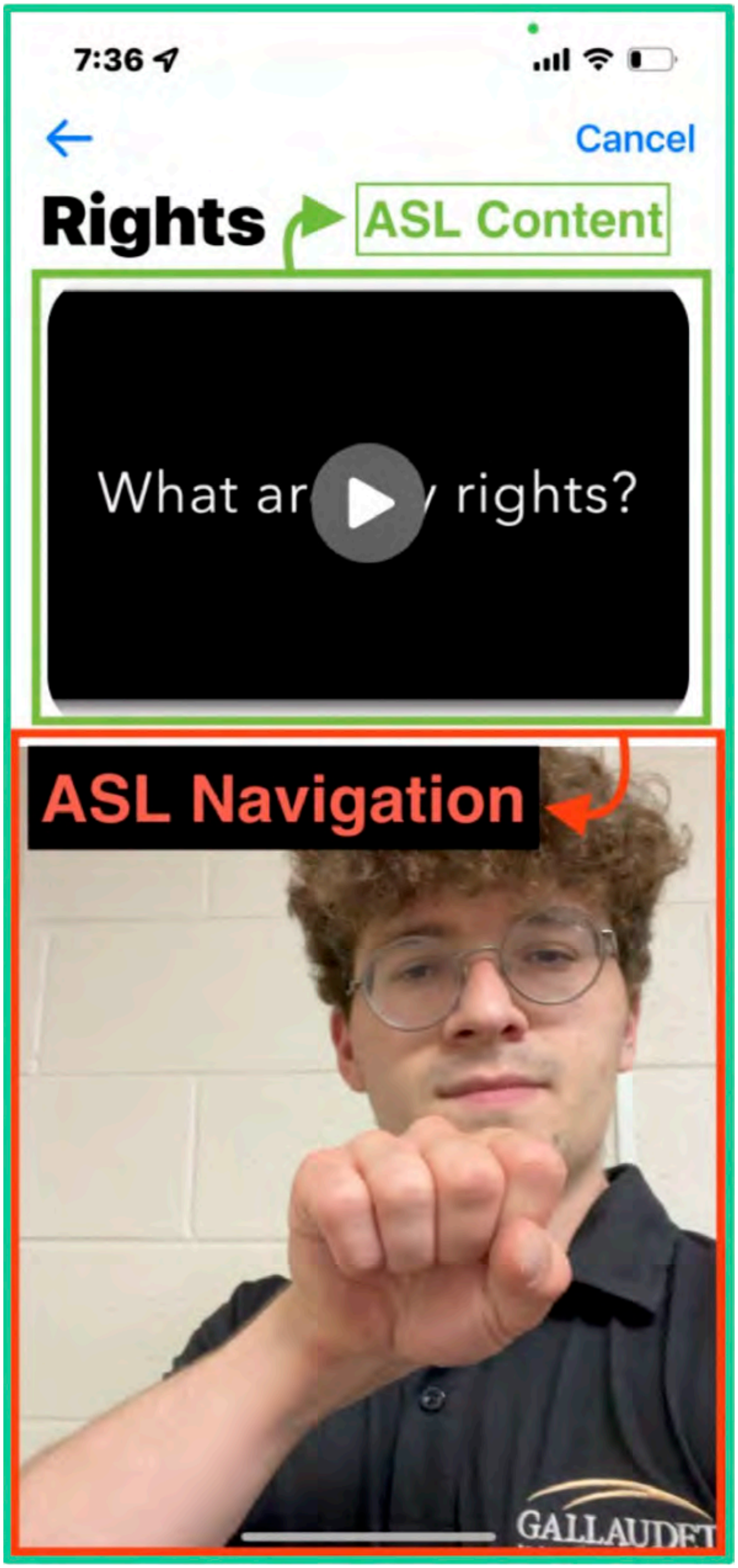


Fig. 3. A labeled ASL section in the ASL-Consent app with ASL content and ASL navigation.

Fig. 4. A labeled signature section in the ASL-Consent app with *ASL content* and *ASL navigation.*

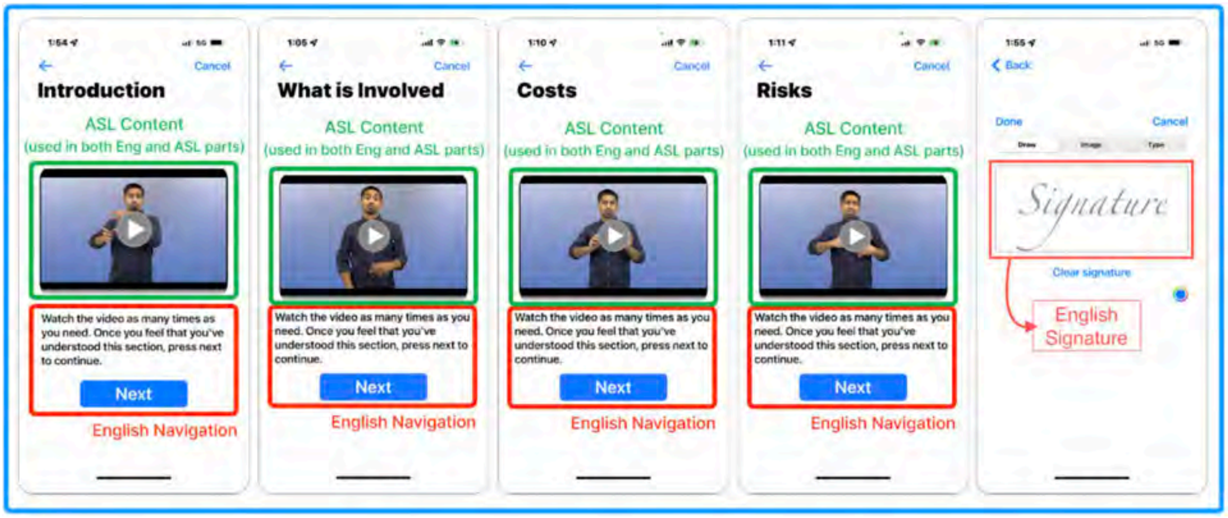


Fig. 5. The English Version of the ASL-Consent app.

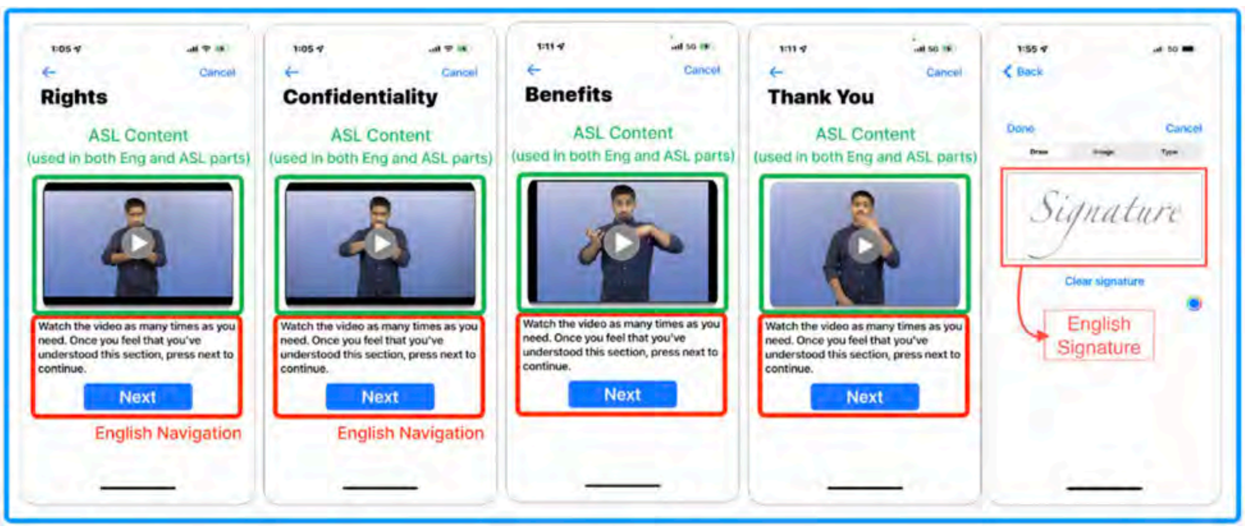


Fig. 6. The ASL Version of the ASL-Consent app.

Discussion

Methodology

To test whether making the digital informed consent process interactable in ASL would improve its accessibility for deaf participants, we developed an iOS mobile app that presented two different versions of the informed consent process to compare two conditions: *English* vs. *ASL navigation* and *English* vs. *ASL signature*. Both informed consent processes were made up of 8 sections of ASL content videos and a signature page, but the *English Version* (Fig. 5) required the participant to navigate from one section to the next through a traditional "English button" and provide a traditionally English "pen-and-paper" signature. The *ASL Version* (Fig. 6) on the other hand, required the participant to navigate from one section to the next by signing "yes' in front of their camera and providing their signature by fingerspelling their full name in ASL. The *ASL Version* also had instructional videos that would play after each ASL content video and explain in ASL how to navigate to the next section using "ASL navigation" and sign the digital consent form with fingerspelling.

With the app, we performed a user experience study with 15 deaf participants who were at least 18 years or older and whose primary language was American Sign Language. With each participant, we tested the two conditions through a one-hour A/B testing session over Zoom where one half of participants were shown the English version first and ASL version second and the other half of participants were shown the ASL version first and English version second.

After showing both the English and ASL versions, participants were asked to respond to nine different standardized usability questions through a 5-point Likert scale. After answering these questions, participants were asked to share any feedback they had about the English and ASL versions or feedback they had about the app in general. For their time, each participant was compensated $25.

Results

Demographics

Ten of the participants self-identified as female, 5 self-identified as male, and 1 self-identified as non-binary. Age ranged from 33 to 77 years of age (mean age = 36 years; SD = 1.41). 10 participants self-identified as White; 3 self-identified as Asian; and 3 self-identified as

Black. Most of our participants were college educated (13% with some college; 63% with a college degree; 25% with a postgraduate degree). As for the geographical distribution of our participants across the United States, 13% of participants were from the East, 19% were from the Midwest, 13% were from the Northeast, 25% were from the South, 6% were from the Southwest, and 25% were from the West.

Quantitative Results

Fig. 7 summarizes the responses participants gave by showing the mean average response for each of the nine user experience questions. To compare the difference in how users felt about the *English part* and *ASL part*, we performed a T-Test on each question to see if there was significant difference. For each question's T-Test, it compared the means between the responses for the "English-system" and the "ASL-system." *Table 1* includes each question and summaries the results for each of their T-Tests.

Table 1. Quantitative responses from our 9 user experience questions regarding participant perceived experience navigating and signaturing in the *English* and *ASL parts.*

* Each question was asked in two separate parts (a) and (b) for the English part and ASL part and the participant gave an answer for both parts separately.

Evaluation Question*	**Summary of Responses**
1. I found the (a) English-system (b) ASL-system unnecessarily complex.	Most people said both the English section and ASL Sections were not complicated or had little-to-no complication.
2. (a) I feel comfortable using the (a) English (b) ASL system.	Most participants felt that both English and ASL navigation and signature were "easy-to-use".
3. I thought the (a) English-system (b) ASL-system was easy to use.	Most participants felt that both English and ASL navigation and signature were "easy-to-use".
4. It was easy to learn to use the (a) English-system (b) ASL-system.	Most participants said that both versions felt "fairly easy" or "easy" to learn.

Evaluation Question*	**Summary of Responses**
5. Using the (a) English-system (b) ASL-system was a frustrating experience.	Most participants said that both versions were “barely frustrating” or “not frustrating at all.”
6. It was easy to provide my signature at the end of the (a) English-system (b) ASL-system.	Most participants said that both versions were easily signaturable.
7. Overall, I am NOT satisfied with the amount of time it took to navigate to the next section in the (a) English-section (b) ASL-section.	Most participants said that they felt that English and ASL navigation ranged from feeling like a “medium length of time” to “fairly short” or “very short.”
8. Overall, I am NOT satisfied with the amount of time it took to provide my signature at the end of the (a) English-system (b) ASL-system.	Most participants said that they felt both the English and ASL signature was “fairly short” or “very short.”
9. The (a) English-system (b) ASL-system has all the function and capabilities I expect it to have	Overall, participants felt the ASL navigation and signature met their needs and expectations more than the English navigation and signature.

No Significant Difference

After calculating all the p-values for each question one through 9, none of which had significant difference, we found there is no significant difference in the user experience between the *English part* and *ASL part*.

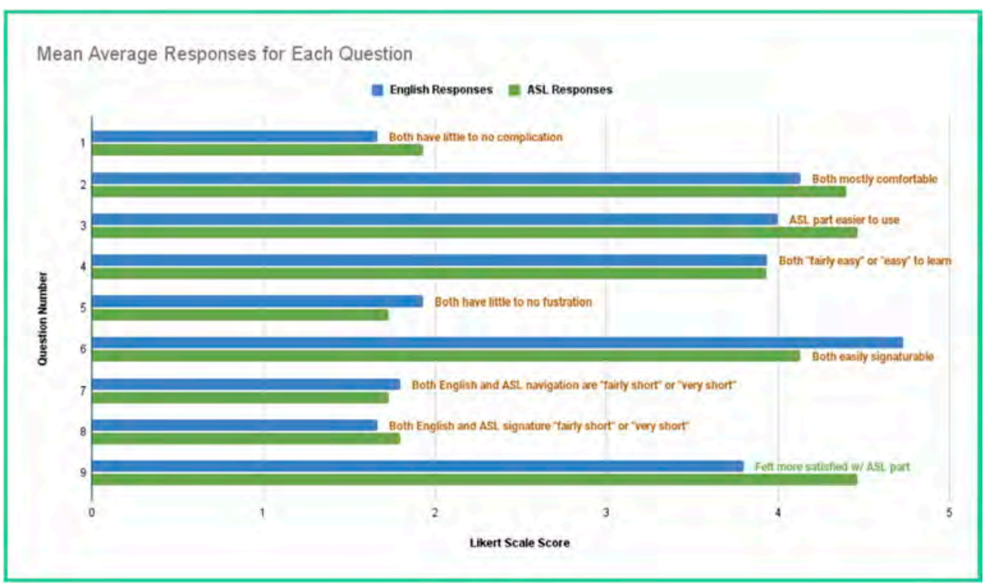


Fig. 7. The mean average of the responses for each of the 9 quantitative questions.

Qualitative Feedback

Overall, participants felt that *ASL navigation* and *signature* felt overall equivalent to or better than their English counterparts. Most participants also openly expressed that they were excited about the general concept of *ASL navigation* and *signature* and felt it was "worth it for deaf people who use ASL."

For the ASL signature section however, participants noted that spelling your full name in ASL felt tedious, especially for those with long names, and felt signing their initials or "I consent" would be faster. One of the biggest concerns amongst participants was that ASL navigation and signature are totally new concepts and will present a learning curve regardless of how much deaf signers like them in the long run. ASL navigation and signature will need to be implemented with this in mind, making good use of an instructional or onboarding process. In line with satisfying the entire community that self-identifies as deaf, deafblind, or hard of hearing, which has a wide diversity of language backgrounds, many participants also suggested giving users the option to pick between traditional English and ASL navigation and signature as well as the option to pick different singing styles for the ASL instructional videos (i.e., Black ASL, West Coast ASL, East Coast ASL, slower signing, faster signing, etc.).

Conclusion

The quantitative responses and qualitative responses show that the deaf participants are in agreement that *ASL navigation* and *signature* in an informed consent app are at least equal to the traditional *English navigation* and *signature* that they are used to. However, we noticed that participants really liked the concept of *ASL navigation* and *signature* but didn't like the way that we implemented it as much. Participants felt that having the ASL content videos only watchable in a full screen mode that took you out of the ASL section screen interrupted the flow of the *ASL* *navigation*, as it forced the user to have to hit “x” in the top left corner to go back to the section screen before they could respond with “yes.” Participants also felt this flow was interrupted by having the same ASL instructions play at the end of each section. Instead of having the ASL instructional videos appear right after each ASL content video, one participant recommended showing it only once before the user starts the informed consent process. If the user ever needed to reference the instructions, they could click on a small button in the corner of the screen to view that section.

The *ASL signature* section also suffered from a similar issue. The ASL instructions for the section need to be improved, along with how the process itself is done. Many participants felt that spelling out their entire name was too tedious and recommended it would be better to let the user fingerspell only their initials or “I consent.”

Regardless, it was clear that participants really liked being able to sign “yes” and being able to automatically move to the next section.

Limitations

As with any piece of research, this study had many limitations. The sample used for this study was very small (only 15 people) and wasn’t very diverse. Most participants had at least obtained a college degree (88%) and thus represented only 18% of deaf people in the U.S (Garberoglio, Cawthon, & Sales, 2019). This is even more significant when considering a deaf individual who grew up using bilingual ASL and English tends to be much more comfortable and fluent in English than an individual who grew up using primarily ASL with less access to English.

There were also many limitations to having the app evaluation interviews over Zoom. As we couldn't put the ASL-Consent app onto the app store for participants to download and install on their own devices, they couldn't test the app for themselves. So, all their responses and feedback have been based on the user experience they had to assume they would have had if they had tested the app themselves. There were also times where technical bugs would plague our interview sessions and force our principal interviewer to stop and restart the app in the middle of the live demonstration. Zoom itself had some limitations as well, as there were times when the participant's screen wasn't big enough to see everything that was being demonstrated.

Future Work

Using the feedback collected from this study, we hope to further develop and improve the ASL-Consent app for a second study with a larger, more diverse group of deaf participants. It is especially important that the group is diverse in its language and education backgrounds, with emphasis on deaf people who use English as a second language. This subgroup is more likely to be marginalized from research participation that requires ability to read and provide consent in English. This second study would also need to be done in a way where the participant can test the ASL-Consent app for themselves. This means app evaluation interviews would either need to be done in-person with the app downloaded and installed on the principal interviewer's phone, or over Zoom with the app downloadable from the app store.

To address feedback from participants, the app used for the second study will replace the ASL instructional videos with a short onboarding process that will take place before the participant starts the informed consent process. The app will also include an updated UI that helps ASL navigation and signature flow better, a simpler way to provide your signature in ASL, and accessibility for participants who self-identify as deaf-blind or low vision.

After the completion of this study, and after the benefits of an app-based informed consent process with fully-ASL *content* and *consent* have been validated, we hope to develop an app-based, accessible informed consent toolkit that researchers nationwide can use to make their app-based studies more accessible to deaf people who use ASL. We won't be making this toolkit from scratch however, as many researchers already make use of Apple's ResearchKit framework. ResearchKit is an open-source framework that allows researchers to easily make powerful iOS apps for medical research with visual consent flows, real-time dynamic active tasks, and surveys.

For future development of *ASL content* in an informed consent app, feedback from participants strongly suggests that a hybrid approach, or an approach that provides both options: purely English and purely ASL, is the best option to meet the needs of deaf people who use ASL.

Acknowledgments

This work was supported by NSF REU Site Grants (#1757836 and #2150429) awarded to Dr. Raja Kushalnagar, PI and Dr. Christian Vogler, Co-PI. This work was also supported by the National Library of Medicine (NLM) of the National Institutes of Health (G08LM013797 awarded to Dr. Poorna Kushalnagar).

Works Cited

Anderson ML, Riker T, Hakulin S, Meehan J, Gagne K, Higgins T, Stout E, Pici-D'Ottavio E, Cappetta K, Wolf Craig KS. "Deaf ACCESS: Adapting Consent Through Community Engagement and State-of-the-Art Simulation." *J Deaf Stud Deaf Educ*, vol. 25, no. 1, Jan 2020, pp. 115-125, https://doi.org/10.1093/deafed/enz035

Bailey, Nicole, et al. "Evaluation of an American Sign Language co-curricular training for pharmacy students," *Currents in Pharmacy Teaching and Learning*, vo. 13, no. 1, 2021, pp. 68-72, https://doi.org/10.1016/j.cptl.2020.08.002

Berke, Larwan, et al. "Design and Psychometric Evaluation of American Sign Language Translations of Usability Questionnaires." *ACM Transactions on Accessible Computing*, vol. 12, no. 2, June 2019, pp. 1-43, doi: https://doi.org/10.1145/3314205

Boudreault, Patrick, et al. "Multimodal Visual Languages User Interface, M3UI." *Journal on Technology and Persons with Disabilities*, vol. 7, 2011, pp. 172-182, http://hdl.handle.net/10211.3/210399

Brooke, John. "SUS: A quick and dirty usability scale." *Usability Evaluation in Industry*, no. 194, 1996, pp. 189-194

Easterbrooks, S. R., & Beal-Alvarez, J. S. (2012). States' Reading Outcomes of Students Who Are d/Deaf and Hard of Hearing. American Annals of the Deaf, 157(1), 27–40. http://www.jstor.org/stable/26235186

Finstad, Kraig. "The usability metric for user experience." *Interacting with Computers*, vol. 22, no. 5, Sept 2010, pp. 323-327, https://doi.org/10.1016/j.intcom.2010.04.004

Garberoglio, CL, Palmer, J., Cawthon, S. & Sales, A. (2019). “Deaf People and Educational Attainment in the United States: 2017.” *Washington, DC: U.S. Department of Education, Office of Special Education Programs, National Deaf Center on Postsecondary Outcomes.* http://dx.doi.org/10.26153/tsw/10052

James R. Lewis. “IBM computer usability satisfaction questionnaires: Psychometric evaluation and instructions for use.” *International Journal of Human–Computer Interaction*, vol. 7, no. 1, 1995, pp. 57-78, https://doi.org/10.1080/10447319509526110

Kushalnagar P, Reesman J, Holcomb T, Ryan C. Prevalence of Anxiety or Depression Diagnosis in Deaf Adults. J Deaf Stud Deaf Educ. 2019 Oct 1;24(4):378-385. https://doi.org/10.1093/deafed/enz017

Kushalnagar, P. “Health Information National Trends Survey in American Sign Language (HINTS-ASL): Protocol for the Cultural Adaptation and Linguistic Validation of a National Survey.” *JMIR Res Protoc.*, vol. 6, no. 9, Sep 2017 pp. 172. https://doi.org/10.2196/resprot.8067. PMID: 28903891; PMCID: PMC5617902.

Kushalnagar, P et al. “Deaf patient-provider communication and lung cancer screening: Health Information National Trends survey in American Sign Language (HINTS-ASL).” *Patient education and counseling* vol. 101,7 (2018): 1232-1239. https://doi.org/10.1016/j.pec.2018.03.003

Kushalnagar, P. & Miller, C. (2019). Health Disparities among Mid-to-Older Deaf LGBTQ Adults Compared to Mid-to-Older Deaf Non-LGBTQ Adults in the United States. *Health Equity*. https://doi.org/10.1089/heq.2019.0009

Lesch H, Burcher K, Wharton T, Chapple R, Chapple K. Barriers to healthcare services and supports for signing deaf older adults. Rehabil Psychol. 2019 May;64(2):237-244. https://doi.org/10.1037/rep0000252

McEwen E, Anton-Culver H. The medical communication of deaf patients. J Fam Pract. 1988 Mar;26(3):289-91. PMID: 3346631.

Sanfacon, K., Leffers, A., Miller, C., Oliver, S., DeWindt, L., Wagner, K., & Kushalnagar, P. (2020). Cross-sectional analysis of medical conditions in the U.S. deaf transgender community. *Transgender Health.* https://10.1089/trgh.2020.0028

Singleton, J. L., Jones, G., & Hanumantha, S. (2014). Toward ethical research practice with deaf participants. *Journal of empirical research on human research ethics: JERHRE*, *9*(3), 59–66. https://doi.org/10.1177/1556264614540589

Smith, S.R., & Chin, N.P. (2012). Social Determinants of Health in Deaf Communities. https://doi.org/10.5772/38036

Development of a Shoulder-Mounted Tactile Notification System for the Deaf and Hard of Hearing

Yuta Murayama[1], Rito Emura[1], Shunya Tanaka[2], Yukiya Nakai[2], Akihisa Shitara[3], Fumio Yoneyama[1], Yuhki Shiraishi[1]
Tsukuba University of Technology, Japan[1]
Alps Alpine CO., LTD., Japan[2]
University of Tsukuba, Japan[3]
a213101@a.tsukuba-tech.ac.jp, a181002@a.tsukuba-tech.ac.jp, shunya.tanaka@alpsalpine.com, theta-akihisa@digitalnature.slis.tsukuba.ac.jp, yonefumi@a.tsukuba-tech.ac.jp, yuhkis@a.tsukuba-tech.ac.jp

Abstract

In this study, a shoulder-mounted tactile notification system is proposed and developed for the deaf and hard-of-hearing (DHH) to go out safely and secure operation. The vibration detection rate, correct response rate, and reaction time are investigated using four types of vibrations with input voltages of 1.0, 3.0, 5.0, and 7.0 V for 24 DHH people. Additionally, the location and vibration time of oscillators that satisfy the two conditions of "being able to perceive the movement as a single point" and "being able to recognize the direction quickly without getting lost" are investigated for six DHH people. The experimental results reveal that the parameters suitable for tactile presentation are extracted based on objective and subjective surveys.

Keywords

Deaf and hard of hearing, DHH, accessibility, tactile information, phantom sensation, wearable.

Introduction

Numerous deaf and hard-of-hearing (DHH) people (including the elderly) have difficulty distinguishing between sound generation and sound source direction. Therefore, visual or tactile presentation is used to convey this information. In the case of a visual presentation, sound generation is transmitted by light, and the latter is displayed on a smartphone. In the case of a tactile presentation, the vibrations of smartphones and smartwatches are used (Alles, 1970).

For them to go out safely and securely, the warning sounds generation and sound sources direction must be transmitted via a visual or tactile presentation. However, the use of a visual presentation with these devices is burdensome because the user must continue observing the screen. In addition, transmitting the direction of the sound source is challenging. To transmit sound source direction using tactile presentation, a shoulder-mounted wearable device was developed and a basic study was conducted herein. Note that the latter vibrates at the same points as the sound source direction among four points on the left clavicle, left shoulder, right shoulder, and right clavicle.

Moreover, the direction of the sound source may change, depending on the type of sound generated. To transmit such changes, the vibration direction must be changed according to the direction of the sound source. Therefore, a notification system that utilizes mobile phantom sensations was studied (Hidaka et. al., 2021) using the developed device. In deaf culture, “shoulder tapping” is often used instead of voice as a means of calling out to someone (Coates and Sutton-Spence, 2001). Thus, in this study, “shoulder tapping” was focused on and the effectiveness of a vibration pattern like a “shoulder tap” for communication was investigated. A vibrotactile waist belt is an example of a device that transmits vibrations (Van Erp et al., 2005). However, the vibration point is at the waist and cannot reproduce the sensation of being tapped on the shoulder. The system used in this experiment is illustrated in Fig. 1. The worn device is shown in Fig. 2.

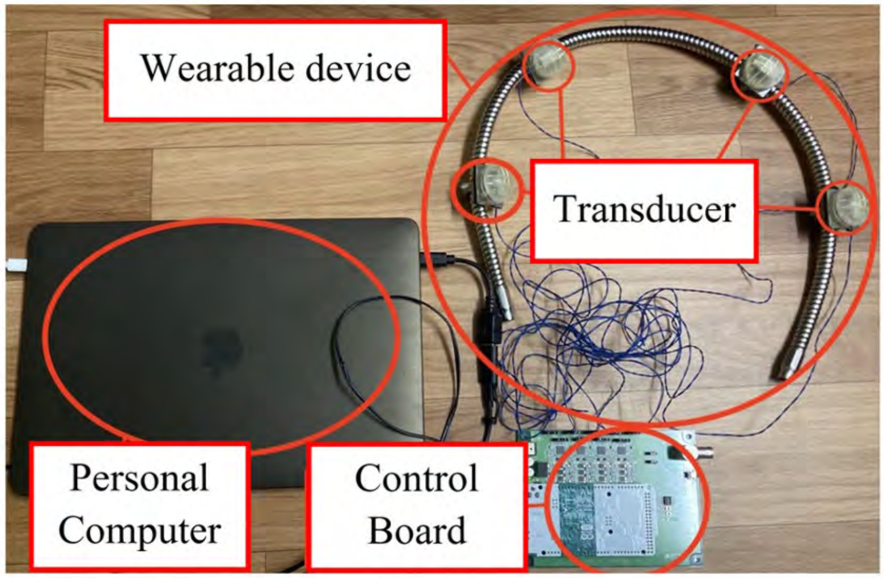


Fig. 1. Experimental system diagram.

Fig. 2. Developed device is worn.

Generally, a personal computer generates a vibration waveform, and the transducer vibrates via a control board. Experiments were conducted to search for a vibration intensity that satisfies the three conditions of: noticing the vibration with certainty, knowing the location of the vibration, and noticing the vibration quickly.

Experiments and Discussion

A transducer was attached to a vibration detection rate, correct response rate, and reaction time using four types of vibrations with input voltages of 1.0, 3.0, 5.0, and 7.0 V for 24 DHH people (68–125 dB hearing level, 20–22 years old). For each type of vibration, the left clavicle, left shoulder, right shoulder, and right clavicle vibrated 20 times at random locations. This was performed for all the four vibration types.

This study was approved by the Research Ethics Review of Tsukuba University of Technology (approval number 2020-11). The duration of the experiment was 90 min, and the honorarium paid to the participants was 1,305 yen.

Table 1 lists the vibration detection and correct response rates for each vibrational strength and vibrational location. Table 2 presents the measured reaction times for each vibrational strength.

Table 1. Detection rate and correct response rate for each vibrational strength.

Detection rate/Correct response rate

Vibrational strength (V)	**Left clavicle**	**Left shoulder**	**Right shoulder**	**Right clavicle**	**Mean**
1.0	0.762/0.770	0.771/0.619	0.762/0.483	0.713/0.494	0.752/0.458
3.0	1.00/0.771	1.00/0.839	0.991/0.853	0.991/0.479	0.996/0.735
5.0	1.00/0.770	1.00/0.936	1.00/0.905	1.00/0.452	1.00/0.731
7.0	1.00/0.721	1.00/0.910	1.00/0.878	1.00/0.484	1.00/0.750
Mean	0.941/0.702	0.943/0.765	0.938/0.756	0.937/0.452	

Table 2. Reaction time for each vibrational strength.

Vibrational strength (V)	**Average reaction time (ms) Mean (SD)**	**Number of data**
1.0	927 (0.381)	0.752/0.458
3.0	693 (0.315)	0.996/0.735
5.0	642 (0.281)	1.00/0.731
7.0	579 (0.241)	1.00/0.750

The experimental results are presented below. Significant differences in the vibration detection rates between 1.0 V and all other combinations were confirmed ($p < .05$, Holm test). For the reaction time, significant differences were observed for all combinations ($p < .05$, Holm test). The vibration detection rates were 100 % at 5.0 and 7.0 V, and the difference in reaction time between 5.0 and 7.0 V was as small as 63 ms on average. By contrast, regarding correct response rates for accurately recognizing the vibration location, significant differences were observed between 1.0 V and all combinations. These results indicate that 5.0 and 7.0 V were equally effective in detecting the onset and direction of the vibrations. A questionnaire administered to the experiment participants also confirmed the need for a system to transmit the direction of the sound source.

In a study utilizing a moving phantom sensation, the location and vibration time of oscillators that satisfy the two conditions of “being able to perceive the movement as a single point” and “being able to recognize the direction quickly without getting lost” were investigated for six DHH people. The transducer locations are shown in Fig. 3.

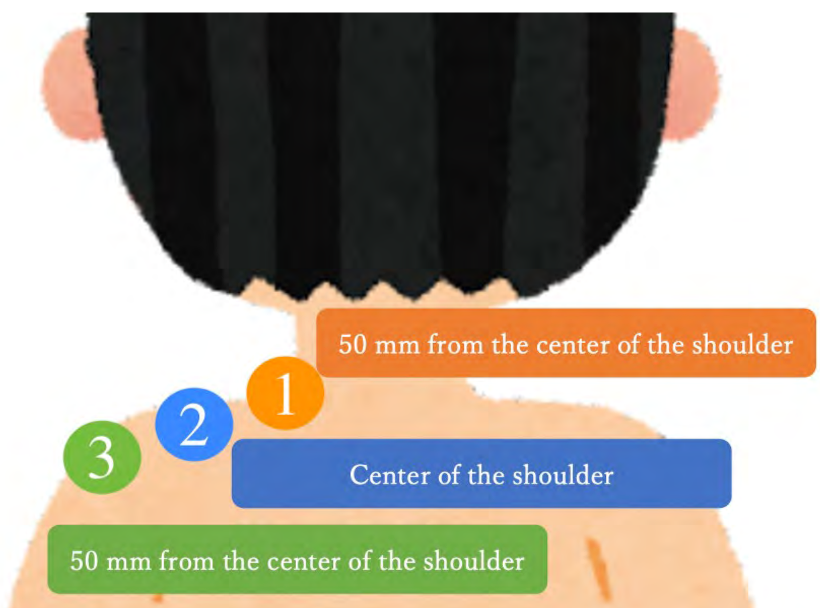


Fig. 3. Vibrator arrangement diagram.

In particular, the transducers were placed at two of three locations: the center of the shoulder (as shown in Fig. 3), in the medial direction from the center of the shoulder, and in the lateral direction from the center of the shoulder. Durations in the range of 0.1–0.5 s were investigated. The transducers were placed 50 mm from the center of the shoulder.

In “Perceiving movement as a single point,” one of 10 different vibrations (direction of movement (shoulder side, neck side) × vibration time (0.1, 0.2, 0.3, 0.4, 0.5 s) was performed at random timing; hence, as soon as the direction of movement of the vibration was known, the button was pressed with the right hand (perceived time measurement of movement direction). The participants were then instructed to input the “direction of movement,” “reliability of response,” and “quality” (an overall evaluation of “feeling the vibration in one place,” “smooth movement,” “vibrating with the same intensity,” and “linear movement”) into the tablet device with their right hand. The right hand was placed on the button to reduce the measurement error in the perceived time of the direction of movement. In addition, to avoid noticing the direction of movement due to the vibration sound, the participants were instructed to remove their hearing aids or cochlear implants if they wore them. The vibration timing was randomly determined within a range of 3–10 s from the start to the next vibration. The experiment was conducted using the same procedure as above, with four types of vibrations: clapping, strong vibration, incoming Apple phone calls, and an iPhone on vibration (please note that iPhone is a trademark of Apple Inc).

After all the vibration presentations, the experimental participants randomly entered one of the four questions on a five-point Likert scale (1: evil, 5: good). Participants randomly entered a 5-point Likert scale (1: poor, 5: good) for the four questions, “A tap on the shoulder,” “Comfort,” “Ease of noticing,” and “Vibration I would like to use when I receive a phone call.”

This study was approved by the Research Ethics Review of Tsukuba University of Technology (approval number 2021-21). The duration of the experiment was 90 min, and the honorarium paid to the participants was 1,305 yen.

Table 3 lists the correct response rates, direction perception time, confidence in direction responses, and quality of continuous movement for each vibration time. The confidence in the direction response reflects the confidence of the respondent in answering the direction of movement of the vibration in 10 % increments from 0 to 100 %. The quality of continuous movement (Rahal, 2009) was evaluated based on the following four items: “Feel the vibration in one place,” “Smooth movement of the vibration point,” “Vibration with the same intensity,” and “Linear movement,” with overall ratings ranging from 0 to 100 % in 10 % increments.

Table 3. Correct response rates, direction perception time, confidence in direction responses, and quality of the continuous movement for each vibration time.

Vibration time (s)	Correct Response rates Mean (SD)	Correct response rates p <. 05	Direction perception time (s) Mean (SD)	Direction perception time (s) p <. 05	Confidence in direction responses Mean (SD)	Confidence in direction responses p <. 05	Quality of continuous movement (%) Mean (SD)	Quality of continuous movement (%) p <. 05
A:0.1	0.630 (0.156)	B, C, D, E	1.01 (0.068)	E	66.4 (3.19)	D, E	52.2 (2.67)	B, C, D, E
B:0.2	0.833 (0.106)	A	1.05 (0.056)	E	72.5 (2.65)		62.9 (2.23)	A
C:0.3	0.898 (0.134)	A	1.11 (0.054)		76.8 (2.53)		68.0 (2.13)	A
D:0.4	0.852 (0.162)	A	1.22 (0.056)		82.0 (2.61)	A	69.0 (2.19)	A
E:0.5	0.889 (0.116)	A	1.32 (0.054)	A, B	79.0 (2.56)	A	67.8 (2.15)	A

The mean correct response rates for the vibration time of 0.2 s and above were significantly higher than that for the vibration time of 0.1 s. The mean percent “Correct” was significantly higher at the vibration time of 0.2 s compared with the vibration time of 0.1 s. The perception time was significantly different for the vibration time of 0.5 s with 0.1 s and 0.2 s combination, with a significantly faster direction perception time below the vibration time of 0.2 s compared with the vibration time of 0.5 s (the vibration time 0.1 s was 0.31 s faster than the vibration time 0.5 s). The mean confidence in direction response was significantly different for the vibration time of 0.1 s with 0.4 s and 0.5 s combination, and significantly higher for 0.4 s and longer. The mean quality was significantly higher for the vibration time of 0.2 s than that for the vibration time of 0.1 s.

Table 4. Correct response rate, direction perception time, confidence in direction responses, and quality of the continuous movement for each vibration location

Vibration location	**Correct Response rates Mean (SD)**	**Correct response rates p <. 05**	**Direction perception time (s) Mean (SD)**	**Direction perception time (s) p <. 05**	**Confidence in direction responses Mean (SD)**	**Confidence in direction responses p <. 05**	**Quality of continuous movement (%) Mean (SD)**	**Quality of continuous movement (%) p <. 05**
A:1-2	0.844 (0.161)	C	1.08 (0.045)	C	74.7 (2.11)	C	64.2 (1.78)	B
B:2-3	0.739 (0.270)	C	1.30 (0.048)	C	69.0 (2.23)	C	57.9 (1.88)	A, C
C:1-3	0.878 (0.162)	A, B	1.03 (0.042)	A, B	82.9 (1.97)	A, B	69.8 (1.66)	B

The mean of the correct response rates in the vibration location 1-3 was significantly higher than that of the others. The mean direction perception time in the vibration location 1-3 was significantly faster than that in the other locations. The mean confidence of the directional responses in the vibration location 1-3 was significantly higher than that of the others. The mean quality confidences in vibration locations 1-3 and 1-2 were significantly higher than those in the other locations. The analysis by vibration duration and location revealed that the best vibration pattern was the vibration time 0.4–0.5 s from the center of the shoulder to the outside. Experiments on the vibration pattern assuming a “shoulder tap” confirmed the system's usefulness. Furthermore, the results of the questionnaire survey confirmed the effectiveness of the developed system.

Conclusions and Future Work

This study focused on the fact that DHH individuals have difficulty identifying the direction of sound generation and require tactile information as a substitute. The initial stage of developing a device that utilizes vibrations and a moving phantom sensation was examined.

Evidently, the fastest and most accurate perception of the onset and direction of vibration was obtained at 5.0 and 7.0 V, with similar reaction times and vibration intensities. Additionally, the clavicle might be unsuitable for transmitting the direction of vibration. Experiments were conducted to find a vibration pattern that satisfies the two conditions of “perceivable as a single point of movement” and “quickly recognizing the direction without hesitation”. Consequently, 0.4–0.5 s (1-3) was found to be optimal, regardless of the direction of vibration presentation. In addition, an experiment to investigate the usefulness of “shoulder tapping” for DHH people when they are called confirmed the usefulness of the “shoulder tap” in the “shoulder-mounted vibration-type notification system”.

In the future, it is envisioned that the system will be used to notify DHH people that someone is calling them out.

Works Cited

Alles, David S. "Information Transmission by Phantom Sensations." *IEEE transactions on man-machine systems*, vol. 11, no. 1, 1970, pp. 85–91.

Coates, Jennifer, and Sutton-Spence, Rachel. "Turn-taking Patterns in Deaf Conversation." *Journal of Sociolinguistics*, vol. 5, no. 4, 2001, pp. 507–529.

Hidaka, Takuma et al. "Pilot Study on Notification Using Phantom Sensation on Hand." *Asian CHI Symposium 2021*, 2021, pp. 157–159.

Legal - Copyright and Trademark Guidelines - Apple. https://www.apple.com/legal/intellectual-property/guidelinesfor3rdparties.html. 2022

Rahal, Lara et al. "Continuous Tactile Perception for Vibrotactile Displays." *2009 IEEE International Workshop on Robotic and Sensors Environments*, 2009, pp. 86–91.

Van Erp, Jan B. F. et al. "Waypoint Navigation with a Vibrotactile Waist Belt." *ACM Transactions on Applied Perception (TAP)*, vol. 2, no. 2, 2005, pp. 10-6117.

Author Index