Volume 9
May 2021
ISSN: 2330-4216

2021

The Journal on Technology and Persons with Disabilities

Scientific/Research Proceedings, Online 2021

CSUN | CENTER ON DISABILITIES

ParaShop: Shopping Assistance to ASD Individuals with a Mobile AR App

Mengting Xia, Nan Chen, Yuemin Tang
The City College, The City University of New York
mxia7498@gmail.com, nanchen432@gmail.com, shirleytang1996@gmail.com
Zhigang Zhu
The City College, The City University of New York and The Graduate Center, The City University of New York
zzhu@ccny.cuny.edu

Abstract

Due to executive functioning challenges typical of Autism Spectrum Disorder (ASD), many people with ASD face barriers with adaptive living skills and developing independence. It is difficult for people with ASD to learn skills such as processing information, social, and speech interaction with the public in a busy environment. Although several studies have used Virtual Reality (VR) to help people with ASD learn daily living skills, users of the VR system may still not be able to finish the real-world tasks, even though they are well trained in the virtual world. In this paper, a mobile AR application for Android and iOS devices, ParaShop, is introduced to help people with ASD go through the supermarket's shopping sequences. ParaShop allows users to create shopping lists with images and automatic categorization of items, which are more intuitive to ASD users. ParaShop also provides instructions with AR scenes to help ASD users to complete their tasks, which is critical for this population. A short YouTube demo video of this application can be found at https://youtu.be/77tp3julNZ0.

Keywords

Autism Spectrum Disorder, Assistive Technology, Augmented Reality, Object Recognition, Text Classification

Introduction

According to the executive summary of the Centers for Disease Control and Prevention (CDC, 2020), Autism Spectrum Disorder is a developmental disability that affects about 1.85% of eight-year-old children in the United States in 2016. Individuals with ASD often have intellectual disabilities, which are deficits in general mental abilities such as planning and problem solving (American Psychiatric Association, 2013). For certain people with ASD, it is challenging to perform daily living skills such as shopping, driving, and personal cleansing independently. It is particularly difficult for them to complete the shopping sequences such as picking up the basket, finding items, finishing payments, and in general communicating with the public in the supermarket. According to the World Health Organization (WHO), ASD cannot be cured. People with ASD require lifelong care and support, which could be costly for their families.

Assistive technologies such as virtual reality (VR) and augmented reality (AR) have been used to support people with ASD. VR offers a complete computer-simulated 3D environment for interaction without the physical world, while AR provides a composite view with the real-world environment and computer-generated images (Chavan, 2016). AR can enhance users' shopping experience in a real environment with the augmentation of virtual annotations.

In this paper, we introduce a mobile AR application named ***ParaShop*** for both Android and iOS devices that can provide step-by-step instructions for the shopping process to individuals who are diagnosed with Autism Spectrum Disorder and have communication difficulties in the supermarket. The key contributions in this work include:

1. ParaShop allows users to create shopping lists with images and automatic categorization of items, which are more intuitive to ASD users.

2. ParaShop provides instructions with AR scenes to assist ASD users to complete their tasks, which is critical for this population.

Related Work

Few studies met the systematic review criteria that utilized the VR system to improve autistic people's daily living skills. One of the studies introduced a VR-based driving system to teach driving skills to adolescents with ASD (Zhang et al., 2017). Some studies are focused on improving the shopping skills of people with ASD. One of the studies used VR to provide real shopping scenarios to help children with ASD (Adjorlu et al., 2017). Another study used a virtual supermarket to test VR training's effectiveness among adolescents with ASD (Lamash et al., 2017). However, these VR scenarios may be very different from the real-world environment. Several teams are developing or have developed indoor shopping AR for navigation. For example, Dent Reality provides indoor AR navigation for shopping malls (Hart, 2019). However, it is not designed for people with ASD and thus does not provide any instructions for people with ASD going through shopping sequences.

Methods and Results

This section will discuss the difficulties that people with ASD might have, design solutions to their problems, and the initial research that we conducted in Goodwill.

User Need Study

Goodwill NY/NJ is a non-profit organization that empowers people with disabilities through employment. Most participants in Goodwill Center at Harlem have Autism Spectrum Disorder. Based on the interviews with service providers and staff in the center, we found several problems participants have in shopping. Our focus in this paper is to help them with the shopping process in the supermarket. According to the service providers of Goodwill Center at Harlem, it is difficult for people with ASD to find out their next steps in a new environment, such as a supermarket, without someone telling them what to do. They usually need a schedule every day with activities including cooking class listed for them to follow. Our app will provide specific instructions for users to follow in shopping. By interviewing staff in the center, when coming to wait in line for checkout, people with ASD do not feel the need to do it. It is also difficult for them to continue their tasks without knowing the time left to leave the place. To solve this problem, our app will show the completion status. In some instances, they leave without knowing to take the change at the end of checkout. This app will provide prompts to remind users to take their change back. People with ASD that struggle in communication cannot ask for assistance in locating items in the supermarket. This app will assist them in finding the correct items on their grocery lists. The steps of shopping are most challenging for people with ASD to complete. Our app is designed to solve all these problems.

Design and Development

Our AR application is implemented to help people with ASD learn and understand how to finish shopping tasks in sequences. The significant features of ParaShop include shopping list, automatic categorization, AR instructions, barcode scanner, and fruit and vegetable recognition.

Technology Stack

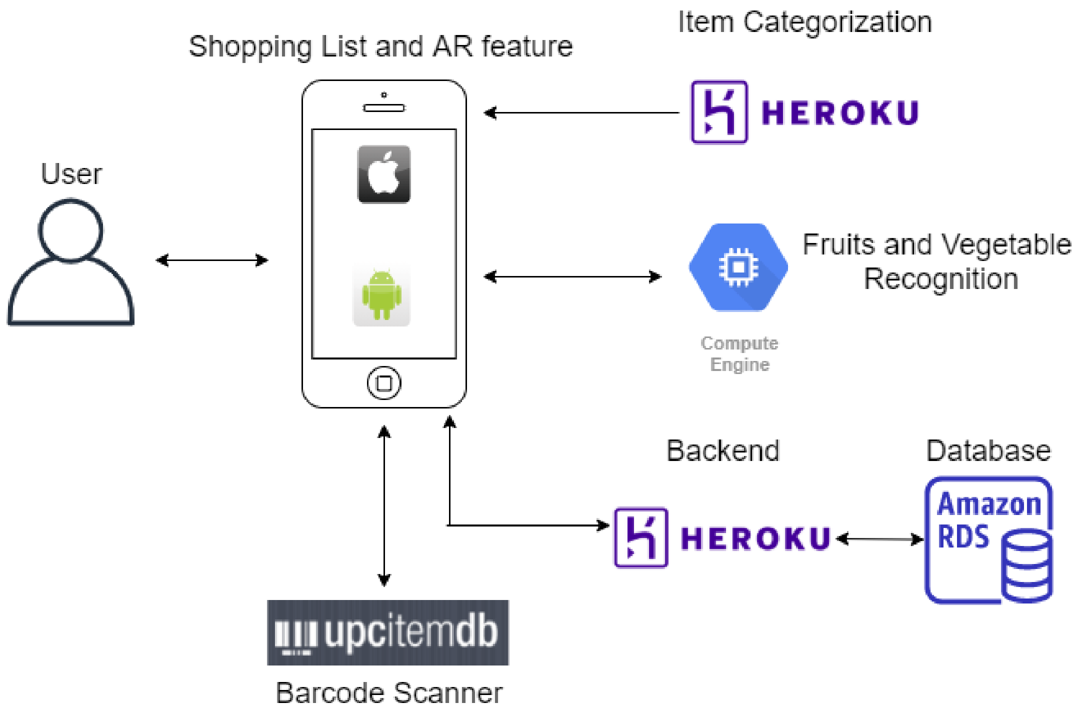

Fig. 1. Technology Stack of ParaShop.

Fig. 1 shows the technology stack of ParaShop. Since our app aims to be cross-platform, we chose to use React Native as the mobile application framework. By using React Native, the frontend of both Android and iOS platforms can be written in one single code base in JavaScript. We used Node.js for the backend and ViroReact for the AR features of our app. The backend is hosted on Heroku and is a RESTful API that accepts all user data. It communicates with our PostgreSQL database, which is hosted on Amazon Relational Database Service (RDS). The text classification model for automatic categorization is hosted on Heroku. The fruit and vegetable recognition model is hosted on Google Cloud Compute Engine. Both recognition models are deployed as Python Flask APIs. Although the fruit and vegetable recognition model is only about 175 MB, the Python Flask API's total size with the Tensorflow package exceeds Heroku's storage limit. Therefore, the fruit and vegetable recognition model is hosted on Google Cloud Compute Engine. UPCitemdb API is used to access barcode information.

Shopping List and Automatic Categorization

Our mobile AR app allows users to create, edit, delete, and share shopping lists (Fig. 2 Left and Middle). The sharing feature allows service providers to share shopping lists with multiple ASD participants. Therefore, they do not need to type the same list for every participant. ASD users can create, edit, and delete items but cannot share them with another user, to avoid confusion to the users. To make our app more user-friendly for people with ASD, we also implement an item selection feature for them using images. With this feature implemented, users do not need to type the item's name when adding the item to the shopping list. Users can tap on the images of the items they want to add to the shopping list (Fig. 2 Right). Currently, there are three categories available in image selection: Beverages, Dairy, and Produce.

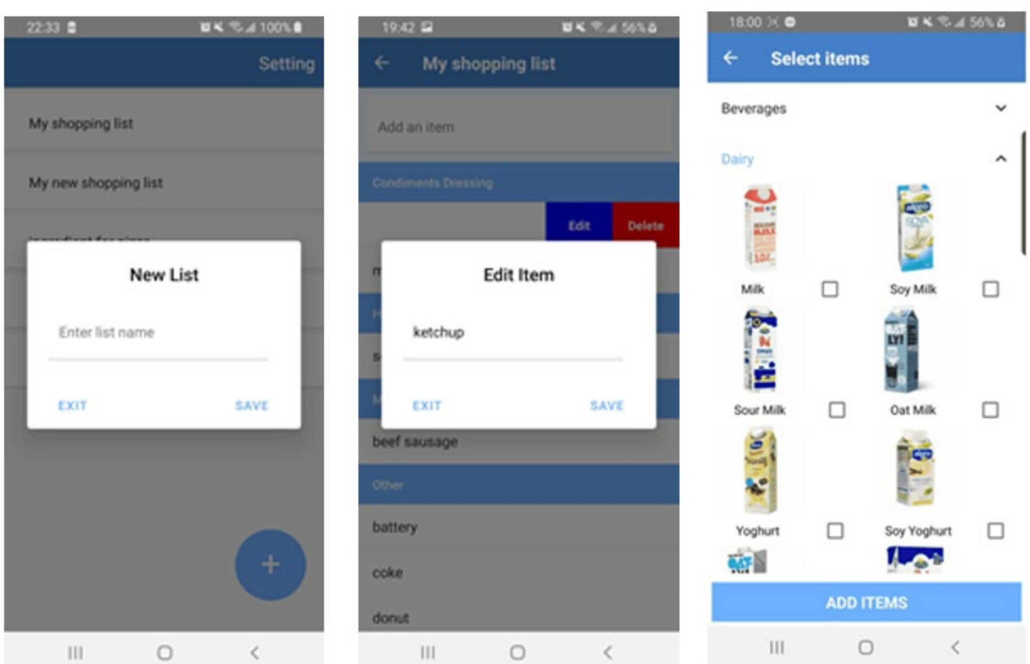

Fig. 2. A Screen for Creating New List (Left); a Shopping List Screen for Editing Item (Middle); and a Screen to Add Items to the List by Selecting Images (Right).

After users add the item to the shopping list, the item is automatically categorized. This helps users speed up the shopping process by finding every item in the same section of the supermarket. The automatic categorization of items involves a text classification model that uses Multi-Layer Perceptron (MLP) Classifier. We found a dataset containing most of the grocery names from an Open Grocery Database Project online (Grocery.com, 2017). The dataset contains the following information of a grocery: group id, upc14, upc12, brand, and the name. Most of the grocery names contain ambiguous abbreviations of words. All the data are preprocessed before input to the MLP Classifier model. Using Python, we dropped all the columns except the name column, renamed all the ambiguous words, and labeled all the items with a category. In the cleaned-up dataset, there are 19 categories, and anything that does not belong to these categories is labeled as 'other'. The 19 categories are *baby, bakery, beverages, breakfast_cereal, confiments_dressing, cooking_baking, dairy, deli, frozen_food, grains_pasta_sides, health_personal_care, household_cleaning, meat, pet_supplies, produce, seafood, snacks, soups_canned_food, and wine_beer_spirits*. Before data preprocessing, there are 110,437 items in the dataset. After data cleaning, only 30,583 items remained for training. We fine-tuned a pre-trained model (Bhatia, 2020) with this training dataset, and the accuracy is about 98.2%.

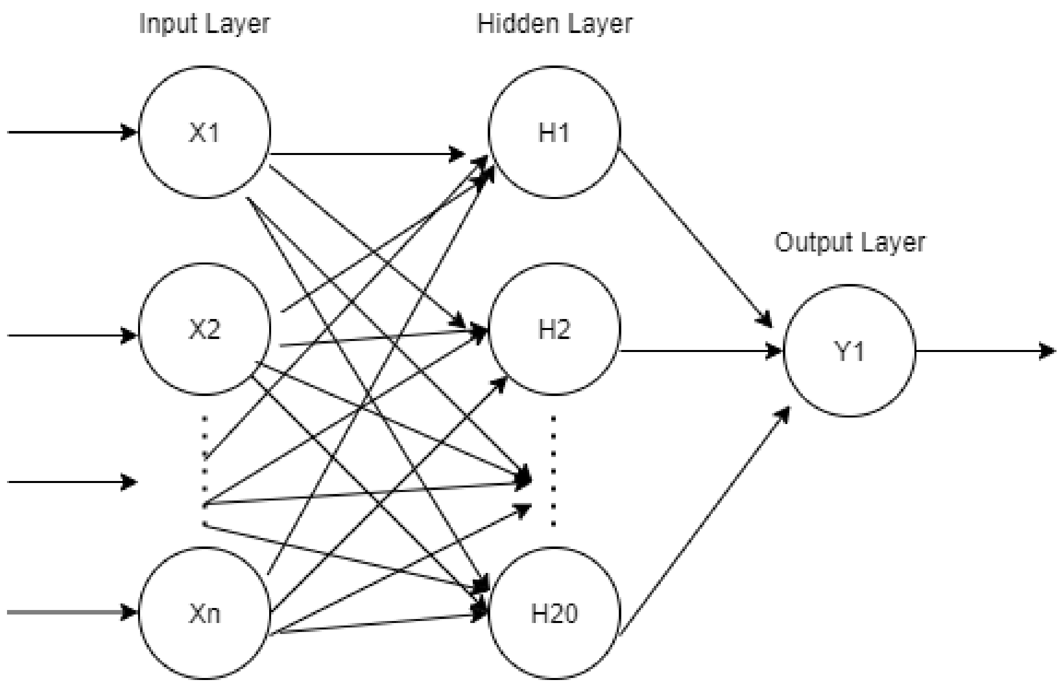

Fig. 3. Item Categorization Model Diagram: The Input is the Name and the Output is the Category.

Fig. 3 shows the structure of the item categorization model. The input layer takes the vectorized feature of a grocery name, which has a length of 6,887. The input length is the number of features extracted from the names using Term Frequency times Inverse Document Frequency (TF-IDF). One unique word in the dictionary of names represents one feature. Then it is mapped to the hidden layer, which has 20 nodes (19 categories + 1 “other” class). Then the hidden layer is mapped to the output layer, which outputs one category for the grocery name.

Step by Step Instruction with Augmented Reality

When people with ASD are exposed to a new environment such as a supermarket, they become anxious because they do not know what to do next (HHS, 2020). Step by step instructions with Augmented Reality (AR) integration are provided for users to follow. The shopping list should be filled before the user goes to the supermarket. In the supermarket, users can open our app and tap on one of the shopping lists they are going to use. In the shopping list screen, users can tap on the “Start” button to start their shopping process. This button will navigate users to the AR scenes.

The real-world scenes of the supermarket will be shown at the top and the instructions will be shown at the bottom of the screen. Users follow the instructions to pick up a shopping cart or basket, find the item based on the ordered list, wait in line, and finally checkout at the end.

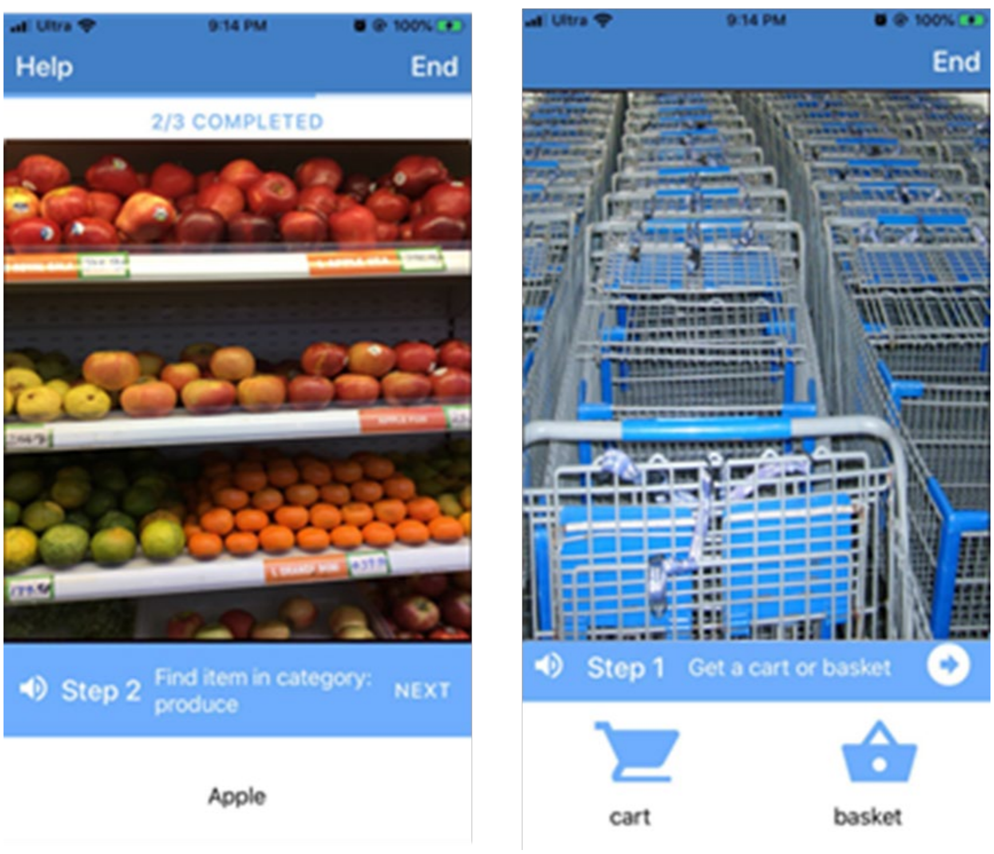

Fig. 4. AR Screen for Finding Apple in the List (Left).

AR Screen of Picking a Cart or Basket (Right).

Fig. 4 shows two of the AR instructions in shopping. When users finish one instruction, they press on the arrow button to go to the next step. So, users no longer need to be confused about what to do next. The item can be skipped if the user cannot find it in the supermarket. Users can choose to skip to the next item anytime. Completion status is shown at the top of the screen for users to know the number of items remaining to be found. This allows people with ASD to know the amount of time left to complete their tasks. During the checkout process, our app would let users choose to pay in card or cash. Each method has its instruction. For better understanding, our application also provides simple voice instructions if the user has difficulties reading texts.

Fruit and Vegetable Recognition

People with ASD also have difficulties recognizing fruits and vegetables with similar or odd shapes. They might not be able to differentiate between apples and pears. To solve this problem, we added the fruit and vegetable recognition feature. Users can use their phone cameras to identify fruits and vegetables through our fruit and vegetable recognition model using Machine Learning (ML). Users take a picture of the item using their phone cameras. Then the label of the item will be shown at the top (Fig. 5 Left). There is a latency of 2 to 3 seconds. This is due to the size of the image and the step it takes for the model to return the label from the cloud. The image user took is resized to 256x256 and converted into a base64 string. Then it is sent to the model hosted on Google Cloud Compute Engine. The string is converted back to the image for model detection in the instance. The model returns an array of labels and sends it to the frontend. The image is deleted from the instance after this process. Users can also access the text-to-speech feature by tapping on the speaker icon (Fig. 5 Left).

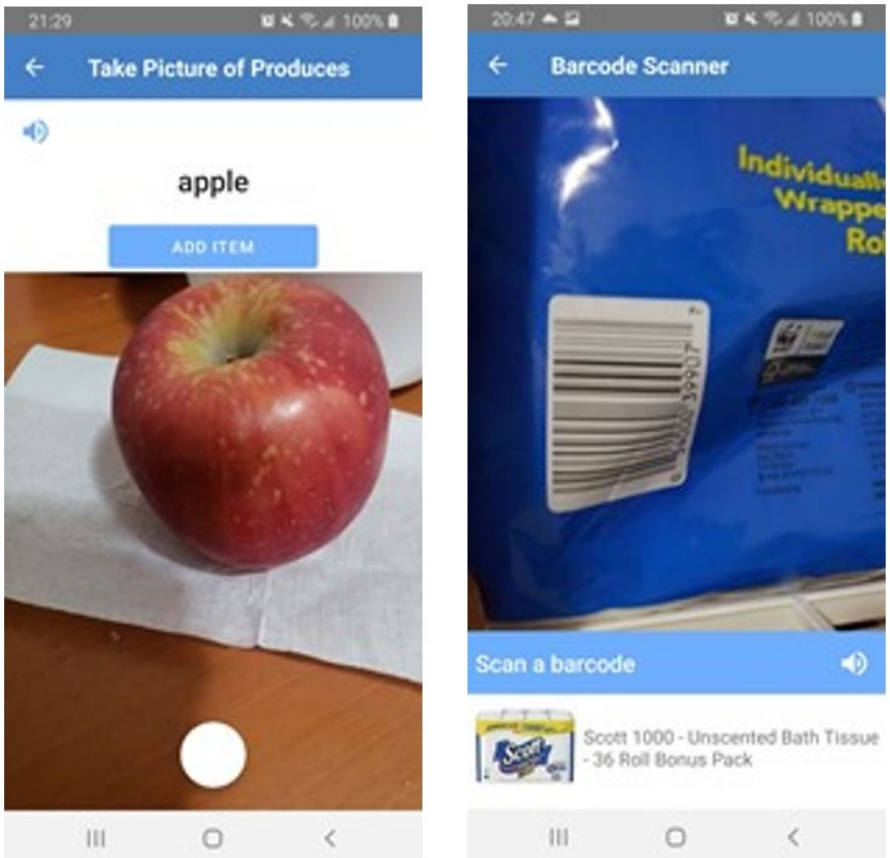

Fig. 5. Fruit and vegetable recognition (Left). Barcode Scanner (Right).

Currently, our model recognizes 13 fruits and vegetables. They are *apple, avocado, banana, carrot, kiwi, lemon, lime, orange, pear, potato, strawberry, tangerine, and tomato*. We have collected 1,181 images of these fruits and vegetables for training our model. The copyright-free images are collected on Pixabay (Pixabay, 2020). Then we used an annotation tool named Labelme to create JSON files for each of the images (Wada, 2020). The JSON file contains the position and label of each item in the image. Figure 6 shows how we labeled each image using Labelme.

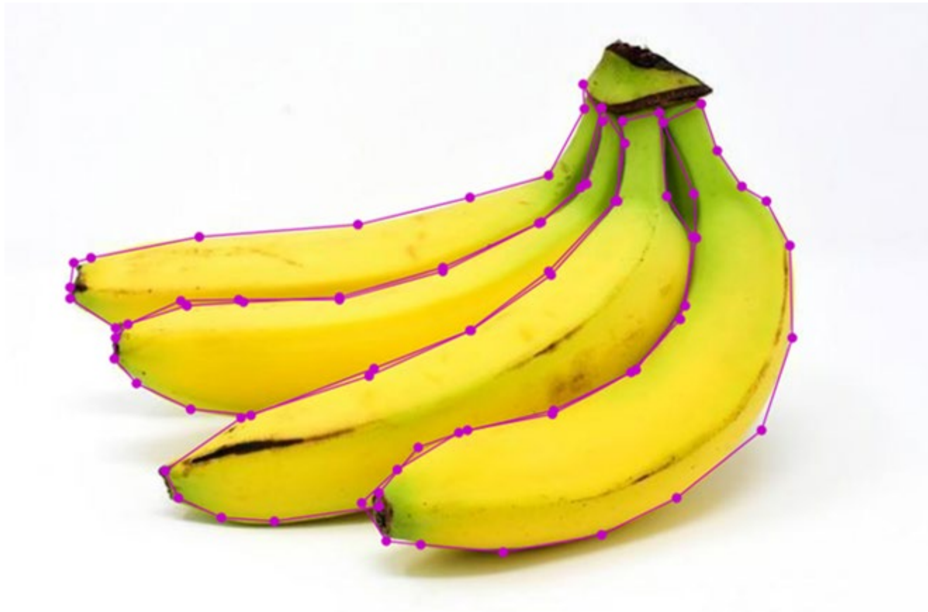

Fig. 6. An Example of Labeling Image Using Labelme.

These labeled images are used during training for our fruit and vegetable recognition model. Since existing models for similar tasks can only recognize three kinds of fruits (Dtcarrot, 2019). It is a Mask Regional Convolutional Neural Network (R-CNN) built on FPN and ResNet101 and was pre-trained with MS COCO dataset.

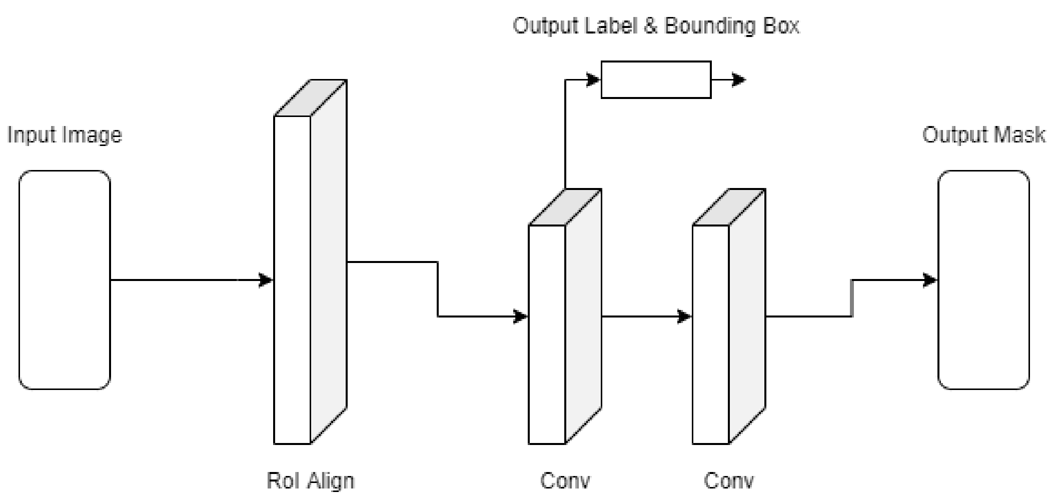

Fig. 7. General Structure of Fruit and Vegetable Recognition Model.

In Fig. 7, there are three outputs: label, bounding box, and mask. Region of Interest (RoI) align layer is used for feature map extraction. The dataset is split into train and test datasets, with 927 images for the train dataset and 257 images for the test dataset. The model has a mean average precision (mAP) of 0.885 at Intersection over Union (IoU) of 0.5 on the test dataset. Intersection over Union is the intersection between the predicted bounding box and the actual bounding box over their union. In this case, a prediction is true positive if IoU is greater than 0.5. The mAP is the average of average precision (AP) over the number of classes.

The model detects objects in the image and then classifies each of the detected objects based on the labels. It takes images of size 256x256 and returns an array of labels, along with the bounding boxes and scores for each of the detected items in the image. As shown in Fig. 8, it detects the fruits in the image and draws out the boundary for each item using a bounding box. The label and the score of each item are shown in the top left corner. However, in the app interface, we set it only to return the labels to make it easier for people with ASD to use the app. It would be hard for people with ASD to understand if the image has all the bounding boxes and scores.

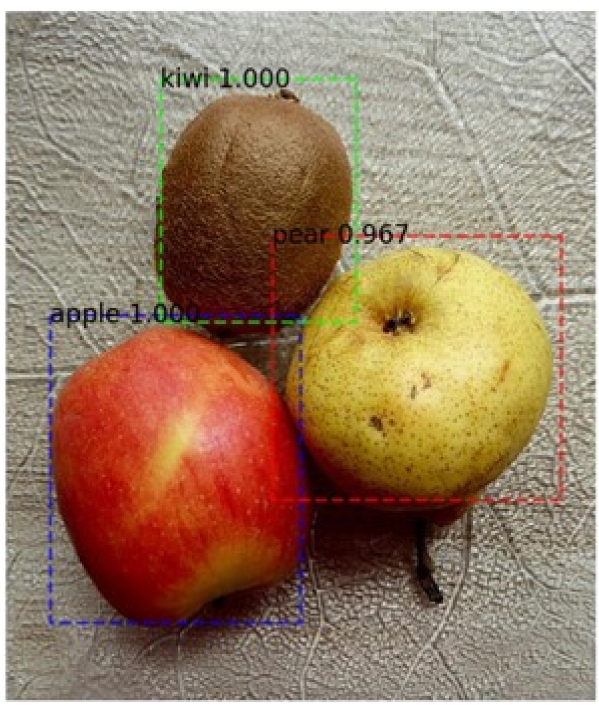

Fig. 8. An Example of the Model Output.

Barcode Scanner

In addition to the fruit and vegetable recognition feature, we also added the barcode scanner to allow users to identify objects other than fruits and vegetables (Fig. 5 Right). We use the npm package "react-native-camera" to extract barcodes in the screen into strings (NPM/CLI, 2020). The string is sent to UPCitemdb API (UPCitemdb, 2020) as an HTTP request to obtain the item's title and image. UPCitemdb API accepts strings for Universal Product Code (UPC) and European Article Number (EAN). UPC and EAN are barcodes widely used for tracking products. The "ADD ITEM" button will only appear if the scanned item is on the shopping list.

User Feedback and Evaluation Plan

By closely connected with Goodwill, we added several new features that would make the app more user-friendly to people with ASD. Based on the feedback from Tasha Stanton, a staff trainer from Goodwill, we added tutorials for barcode scanner and fruit and vegetable recognition features. Also, voice instruction is added for AR instructions and tutorials because users might have difficulties understanding the text. The font size and visual components are designed to be easy to understand for people with ASD.

A video demo of the ParaShop app can be found at https://youtu.be/77tp3julNZ0. A formal user study is planned after the reopening of businesses in NYC, and we have had the IRB approval ready for the tests with ASD users. We planned to have 12 ASD participants to do two similar shopping trips, one with the app and one without, and to compare the performance of various components of the system.

Conclusion and Discussion

ParaShop is a tool to teach shopping sequences to autistic people with help from their family members and service providers. People with ASD can use the tool to go shopping independently and improve their shopping experience.

We are currently adding more fruits and vegetables to our recognition model while keeping the accuracy. In the future, we would like to increase the number of images of fruits and vegetables through the photo users take when they use our recognition feature. It would increase our training dataset and thus increase the accuracy of our model. To make our app more user-friendly, we are planning to allow users to tap on the returned label to highlight specific fruit or vegetable with bounding boxes on the image. We will also work on the image selection feature to provide more images and categories to our users. After the formal user evaluation, we will identify other challenges that people with ASD have and how the app helps them to overcome. The key design principle is to break down each step of implementation with its correlated cognitive ability, e.g., the need to problem solving, the need to navigation, or the need to socially interaction, etc.

Acknowledgments

This work is supported by a CCNY CEN Course Innovation Grant from Moxie Foundation. Additional advice given by Dr. Erik K. Grimmelmann and Dr. Hao Tang have been a great help in the development of fruit and vegetable recognition. We would also like to extend our thanks to Celina Cavalluzzi, Tasha Stanton, Jason Ramirez, their colleagues, and their ASD participants at Goodwill NY/NJ for their guidance and their feedbacks, and Dr. Cecilia Feeley of Rutgers University for her discussions on the need of people with ASD in shopping. An early version of this work was presented at the 2020 IEEE MIT Undergraduate Technology Research Conference, October 9-11, 2020 (Xia et al., 2020).

Work Cited

Adjorlu, Ali, Høeg, Emil, Mangano, Luca, Serafin, Stefania. "Daily Living Skills Training in Virtual Reality to Help Children with Autism Spectrum Disorder in a Real Shopping Scenario." *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, 13 Oct. 2017, doi:10.1109/ismar-adjunct.2017.93.

American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*. American Psychiatric Publishing, 2013.

Bhatia, Akshay. akshaybhatia10. "akshaybhatia10/Book-Genre-Classification." *GitHub*, https://github.com/akshaybhatia10/Book-Genre-Classification. Last visited Oct 11, 2020.

Chavan, S. "Augmented Reality vs. Virtual Reality: Differences and Similarities." (2016).

CDC. "Executive Summary." *Centers for Disease Control and Prevention*, 2 Oct. 2020, www.cdc.gov/ncbddd/autism/addm-community-report/executive-summary.html.

Dtcarrot. "Segmenting Fruits with Mask RCNN." *Kaggle*, Kaggle, 14 July 2019, www.kaggle.com/dtcarrot/segmenting-fruits-with-mask-rcnn.

Grocery.com. "Open Grocery Database Project." *Grocery.com*, 24 Jan. 2017, www.grocery.com/open-grocery-database-project/.

Hart, Andrew. "Dent Reality-Building the Future of Indoor AR Navigation." *Medium*, Dent Reality, 26 Sept. 2019, https://medium.com/dent-reality/dent-reality-building-the-future-of-indoor-ar-navigation-9dd6431b31d6

HHS, "Autism Spectrum Disorder: Communication Problems in Children." *National Institute of Deafness and Other Communication Disorders*, U.S. Department of Health and Human Services, 2 Oct. 2020, www.nidcd.nih.gov/health/autism-spectrum-disorder-communication-problems-children.

Lamash, Liron, Klinger, Evelyne, Josman, Naomi. "Using a Virtual Supermarket to Promote Independent Functioning among Adolescents with Autism Spectrum Disorder." *2017 International Conference on Virtual Rehabilitation (ICVR)*, 15 Aug. 2017, doi:10.1109/icvr.2017.8007467.

Pixabay, "1.8 Million+ Stunning Free Images to Use Anywhere - Pixabay." *Pixabay*, https://pixabay.com/. Last visited Oct 11, 2020

"Npm Javascript Package Manager." *Npm*, https://docs.npmjs.com/cli/v7/commands/npm. Last visited Oct 12, 2020

UPCitemdb, "UPCitemdb API." *UPCitemdb,* https://devs.upcitemdb.com/. Last visited Oct 11, 2020

Wada, Kentaro. "Wkentaro/Labelme." GitHub, https://github.com/wkentaro/labelme. Last visited Nov 23, 2020

WHO, "Autism Spectrum Disorders." *World Health Organization*, World Health Organization, www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders.

Xia, Mengting, Nan Chen, Yuemin Tang, and Zhigang Zhu, "ParaShop: A Mobile AR App in Assisting People with ASD in Shopping." 2020 IEEE MIT Undergraduate Technology Research Conference, October 9-11, 2020.

Zhang, Lian, Wade, Joshua, Bian, Dayi, Fan, Jing, Swanson, Amy, Weitlauf, Amy, Warren, Zachary, Sarkar, Nilanjan. "Cognitive Load Measurement in a Virtual Reality-Based Driving System for Autism Intervention." *IEEE Transactions on Affective Computing*, vol. 8, no. 2, 2017, pp. 176–189., doi:10.1109/taffc.2016.2582490.

The Journal on Technology and Persons with Disabilities

Share&Care: A Family Interaction Application for Older Adults Living Alone

Jin Chen, Yu Ting Yang, Xiaohong Zhu, Zhigang Zhu

Dept. of Computer Science, The City College, The City University of New York

jchen025@citymail.cuny.edu, yyang004@citymail.cuny.edu,

xzhu000@citymail.cuny.edu, zzhu@ccny.cuny.edu

Abstract

Loneliness is a common problem that many older adults experience when they are living alone. Family interaction through telecommunication using mobile applications brings many technical challenges to older adults due to their lack of technological familiarity: many mobile applications are not designed to be senior-friendly. In our work, a user study was conducted on older adults' family interaction and their mobile phone usage through in-person interviews with 12 older adults and an online survey for 28 young people age from 16 to 35. By analyzing the survey results, we propose a senior-friendly communication application, Share&Care, which aims to reduce older adults' loneliness by promoting real-life family interaction and improving their experience with the mobile communication application between them and their family members. The solution integrates senior-friendly user interfaces such as voice navigation, remote family management, and multi-layer security to increase the older adults' comfortability and trust in our application. Interested readers can watch a YouTube video introducing the Share&Care app at https://youtu.be/KeDxeV-5_Rk.

Keywords

Senior-Friendly, User Study, Family Communication, Voice Navigation, Remote Management

Introduction

The population aged 65 and older is growing at an unprecedented rate with the increasing global life expectancy and is expected to increase from 9% of the world population in 2016 to 12% by 2030 (Roberts et al., 2018). Older adults usually have poorer health conditions, such as hypertension, dementia, heart failure, etc. (Marengoni et al., 2008), and as such they will need more attention and care from their families. However, there are many barriers for young people to take care of their parents or grandparents, such as the physical distances of their living places. Over a quarter of the U.S. population over age 65 live alone in 2010 (West et al., 2014).

Many programs had formed to create better life quality for older adults by keep them healthy, active, happy, and stay connected with people. Among them, Active Assisted Living Programme (AAL) is a European funding program supported by 19 member countries and the European Commission (AAL Programme, 2020). It focuses on overcoming three challenges that older adults face: retirement, loneliness, and chronic disease. Since 2008, it had funded over 220 projects and over 10% of the funded projects are already on the markets. Such programs show the governments' care and effort to draw public attention to improve older adults' life quality, which is also our goal of this paper.

Applications have been developed to help monitor older adults' health conditions (Pavlakis et al., 2012; Gao and Koronios, 2010; Sukmana et al., 2019) usually included features like reminders, emergency buttons, and setting up connections to caregivers. These applications can help older adults with their physical conditions, however, research (Cacioppo and Cacioppo, 2014; Jennifer Yeh and Lo, 2014; Taylor et al., 2018) also show that physically living alone often leads to many mental health problems, such as loneliness and depression. There are limited applications designed for caring about older adults' mental health.

A few studies (Banks and William, 2002; Cacioppo et al., 2014) show that animal interaction can effectively provide companionship for lonely people and improve their mental well-being. Robotic pets (e.g., Tombot Puppy and Joy for All) and virtual pet apps (e.g., GeriJoy) were invented to help older adults curb their loneliness. These products help older adults combat loneliness to some degree, but they are usually expensive and only provide artificial interactions that cannot fully assist older adults in curbing the loneliness. Many existing communication applications (e.g., WeChat, Messenger, Google Duo, etc.) provide great functionalities for social interaction. However, these apps are mainly designed for younger users. These app's extensive features increase their user interface complexity and bring many challenges to older adults since many of them are new to mobile applications.

In this paper, we propose Share&Care, a senior-friendly communication application incorporating voice interaction and remote management to enhance the older adults' experience and accessibility with the mobile application and closely connect them with their families. The key contributions in this work include:

(1) A comprehensive user study on older adults' smartphone usage and family interaction from both younger and older generation's perspectives;

(2) A multi-language iOS application that provides a secure and private communication platform for older adults and their families and friends with a senior-friendly user interface; and

(3) A dual user interface design especially with a senior manager feature, which enables family members to remotely manage older adults' accounts and set up notifications with their permission.

Methods and Results

User Study

To understand older adults' thoughts and needs on family communication through mobile applications, we conducted in-person interviews with 12 older adults and 3 staff members from 3 senior centers. We also surveyed 28 younger people (ages between 16 to 35) to collect the younger generation's feelings and thoughts on communication with their older family members.

Older Adults Smart Phone Usage

Out of the 12 older adults we interviewed, 6 of them used smartphones and others used cell phones. From the 28 younger people we surveyed, 50% of their 58 older family members reported owning a smartphone. Out of the older adults who own smartphones, 56.4% use the iOS system and 43.6% use the Android system.

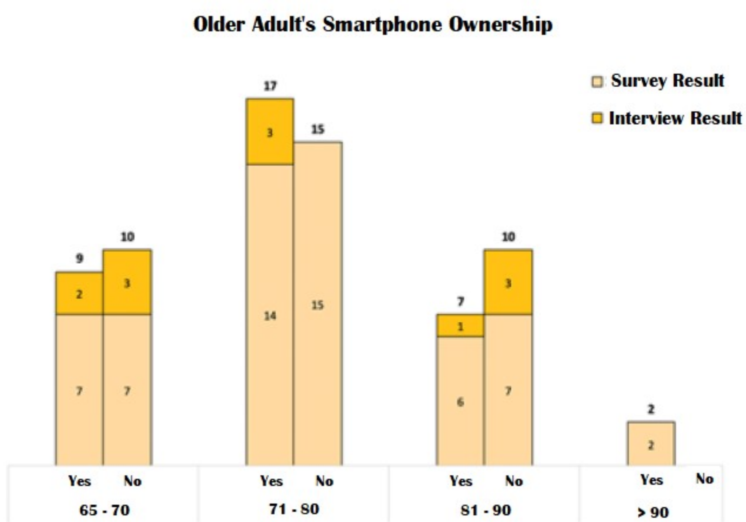

Fig. 1. Older Adults' Smartphone Ownership from the 12 Interviewed Older Adults and 58 Older Family Members of the 28 Survey Respondents.

This data also confirmed by the Pew Research Center study on American mobile phone usage. As shown in "Demographics of Mobile Device Ownership and Adoption in the United States", older adults' smartphone ownership increased from 11% in 2011 to 53% in 2019.

Besides that, older adults' social media usage also increased from 11% in 2010 to 35% in 2015 (Perrin, 2015). With the increase in older adults' smartphone ownership and social media usage, the focus will be on making the applications easily accessible and safe for older adults to use. Many of the older adults we interviewed mentioned that they experienced difficulty in using most mobile applications. Among various issues, such as confusion with button functionalities, hard time with reading small text and clicking on small buttons, and the distraction from irrelevant information, one of their major concerns is personal information security. When older adults register an app account independently, they might accidentally press the wrong button causing unexpected things to pop up and might even request sensitive information (e.g., credit card or ID number). The application complexity, distraction from ads or pop-up, and the risk of losing personal information contributed to the poor user experience and difficulty that older adults have to deal with in many existing applications.

Family Communication Methods

The distance separation between older adults and their family members is a common problem many older adults face, as many younger individuals live in other states or countries away from their older family members for their education or careers. This problem has become severer due to the COVID-19 with strict limitations on travel. Among the 28 younger people we surveyed, 76% of the respondents did not live with their older family members. All the older adults we interviewed either live only with their spouses or alone by themselves.

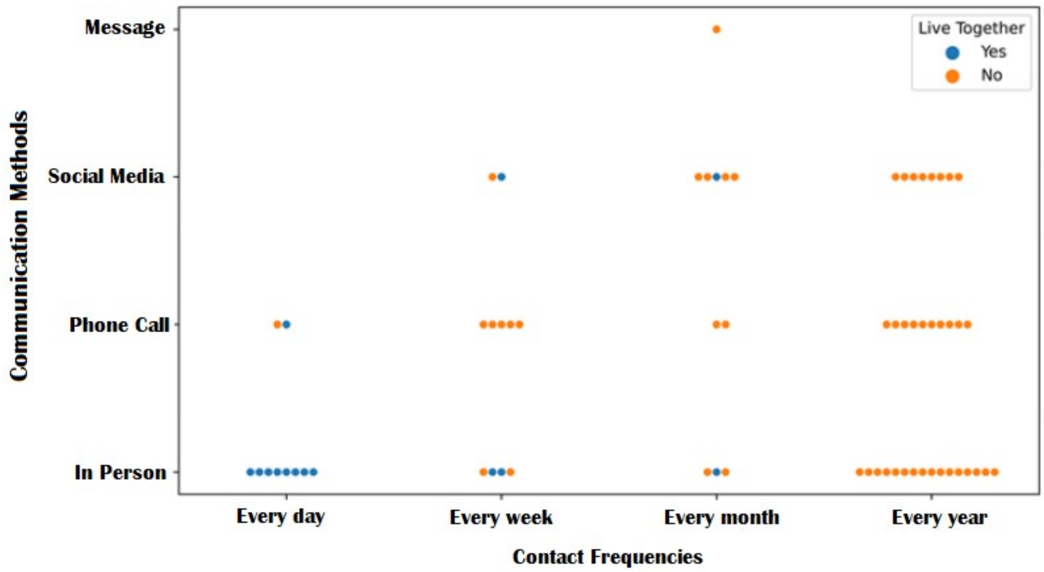

Fig. 2. Young People's Contact Frequencies and Communication Methods with Older Adults.Face-to-face communication and phone calls are the two most common communication methods between older adults and their younger family members. With the distance barrier, young people would communicate with older adults more often if they live with older adults. Over half of the survey respondents only contact their older family members a few times a year (Fig. 2), and this group of people did not live with older adults. However, if older adults can better use smartphones and social media apps, it would allow them to stay closely connected with younger people to better combat their loneliness.

One of the advantages of communicating through social media apps is not requiring immediate responses and it is easy for long-distance communication. Many survey respondents are more likely to increase their communication frequencies with older adults through social media apps. However, most of the interviewed older adults are not actively using social media apps, either due to the lack of access to smartphones or feel that the apps are not personal and distractions from irrelevant information (e.g., promote posts on Facebook). According to the staff members at three senior centers, they all offered technology classes to teach older adults to use computers and smartphones. Older adults are eager to learn new technologies, such as using simple apps to generate videos and share with their families. However, older adults and young people have different needs in mobile applications and most developers often fail to recognize and incorporate older adult's needs when they develop the apps. For example, existing of many small buttons and distraction interfaces due to fancy features that older adults are not going to use, and these ever-increasing app complexities discourage older adults from learning and using the apps.

To minimize these problems, our application focuses on providing a simple and secure communication platform for older adults while maintaining necessary functionalities to engage young people to not only share their lives with older adults but also to help them manage their accounts through a *dual user interface* design – with customized user interfaces to different users and for different functions.

Share&Care Design and Implementation

Share&Care is a multi-language iOS application that provides a secure and private communication platform for older adults and their families and friends with a senior-friendly user interface. It is built with React Native, since React Native is cross-platform and can be easily integrated to serve in both iOS and Android devices. Based on our user study, more older adults are using iOS devices, thus the app is currently developed in the iOS system, but we will expand to Android in the future after more evaluations and tests.

Share&Care provides *basic* communication features as many existing communication applications are offered, including real-time chatting, sharing posts and photos, managing contact lists and user profiles, etc. On top of these basic functionalities, the app also provides *special* features to assist older adults better in interacting with their families and friends. First, to address the older users' major concerns, the app takes special care of users' data security and privacy. Second, the app incorporates voice input and output with a senior-friendly user interface to enhance the older adults' user experience with navigating through the app. Third, the app offers a dual user interface to minimize the unnecessary features for older adults and keep it to attract younger users. The app also offers a multi-language feature to serve users from different countries. Finally, the app allows family members to remotely manage older adults' accounts through permission to free older adults from complex operations and increase family interactions.

Data Security and Privacy

Data privacy is the major concern based on the feedback from the user study. We have found that many older adults don't fully trust mobile applications, either due to massive reports of personal information get misused by applications or experience of personal information get stolen. To ensure user data security and privacy, we add multiple security layers between the app, web services, and databases (Fig. 3). AWS Cognito is used to validate the users' phone numbers and login credentials. Once the login credential is verified, our web service will generate a randomly customized token based on the users' credentials and send to the app. The token is required when the app requests or sends the data to the web service. It will be automatically refreshed, which minimizes the risk of unauthorized access to the web service from external applications.

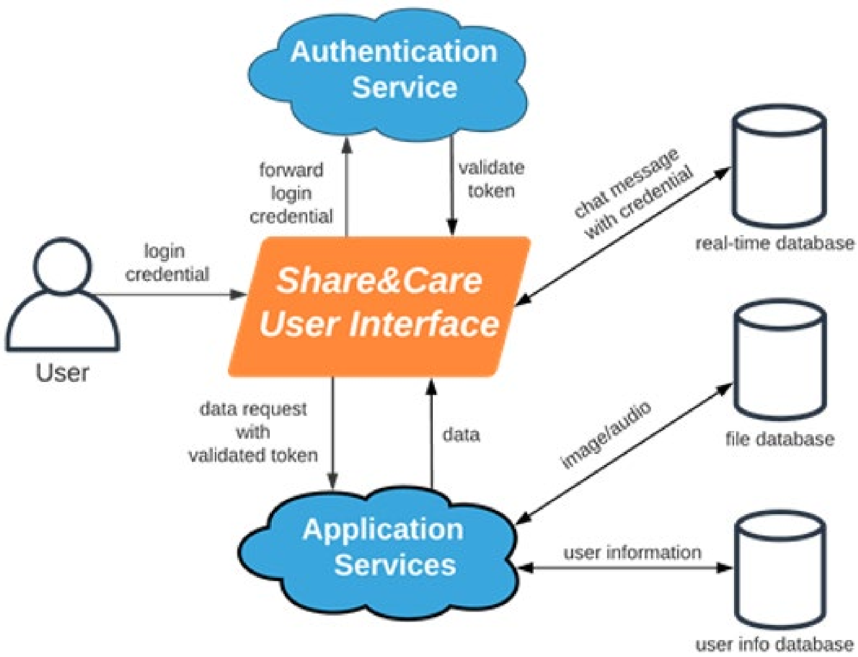

Fig. 3. Share&Care Data Flow Diagram.

To further increase the security of users' data, data are separately stored into three different databases. Firebase is used to handle and store the chat messages for its real-time performance, the app can directly access it with validated credentials. Images and audio files are stored in the Amazon Simple Storage Service (Amazon S3) for its capacity and scalability. Both Firebase and Amazon S3 do not contain users' identities. AWS DynamoDB is used to store user information for its low-latency performance and built-in security; user information will be encrypted through the web service before being stored in the database. Users can successfully register an account only with a valid phone number and the app will not request more sensitive information from older adults. With integrated multi-layer security and multi-databases, our app provides a safe platform for users to be confident in using without worrying about personal information leakage.

Remote Management

Most applications required multiple steps to search and add users, which involves complex commands and the risk of mistakenly adding strangers. Thus, setting up contact lists is a tedious and challenging task for older adults, and they usually need help from younger people, which is difficult as they are living alone. To solve this problem, our app allows older adults to select a family member as their senior manager to remotely manage their accounts with limited authorization. As the privacy concern of the older adults, the senior managers cannot view the older adults' chat messages or directly delete the contact without older adults' permissions (Fig. 4 (a)).

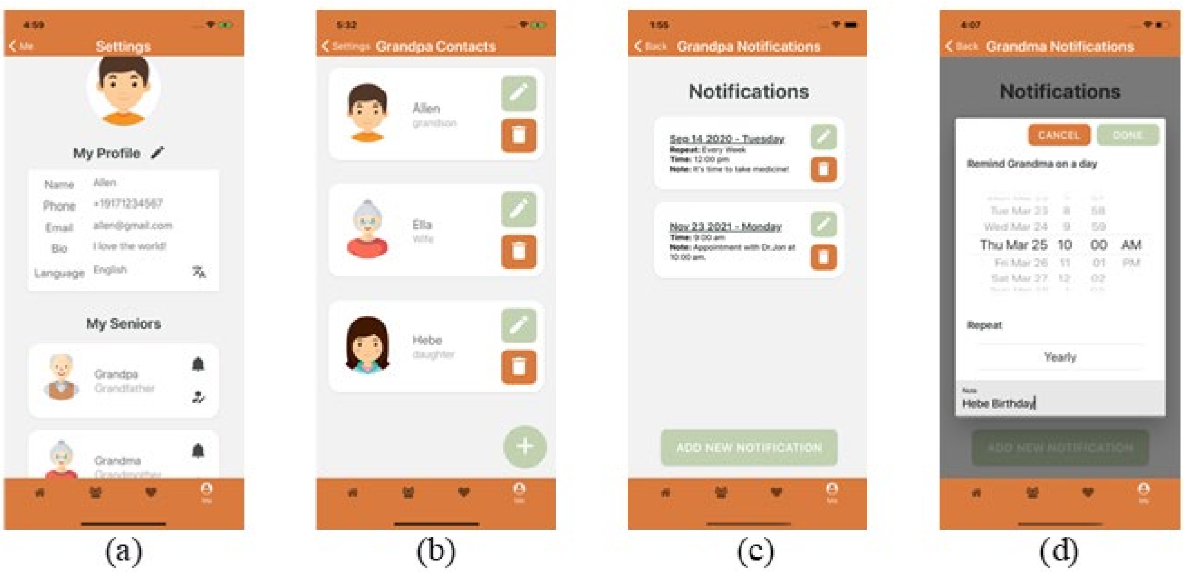

Fig. 4. (a) Senior Manager Setting Screen; (b) Contact Manage Screen; (c) Senior Notification Manage Screen; (d) Set-up Notification Screen.

A senior manager of an older adult can remotely add people to the older adult's contacts, edit the nickname and relationship for each contact, and send a remove contact request that requires approval from the older adult (Fig. 4 (b)). With the contacts set up by the manager, older adults can directly communicate with their families without struggling to search and add people.

Friendly reminders are extremely helpful for older adults, as they are not very good at remembering things. To remind older adults about important events, senior managers can set up notifications for older adults through our app (Fig. 4 (c) and Fig. 4 (d)), such as a doctor's appointment, the time to take a medicine, etc. This notification feature will closely connect the senior manager with the older adults and make older adults feel the care from their families.

Senior-friendly User Interface

Other common problems older adults experience with mobile applications are the size and location of the buttons and relying on younger people for customizing operations (Kurniawan, 2008). Therefore, a simple but unique user interface is essential for older adults to be comfortable and successful in using the application. Our app dramatically simplifies the user interface and incorporates voice interaction to minimize the required screen interaction.

The app is designed with two primary colors and big buttons that can vary by size based on feature priority and large text suggested by Gao and Koronios (2010) to provide better visual aids and reduce distraction. Besides that, voice navigation will be automatically turned on when the older adult logs into the app, to help older adults who have a hard time controlling the screen and allow them to use voice command to navigate through the family posts and photos on the home screen (Fig. 5 (a)). Voice feedback will be provided to notify them the actions will be performed, so it will not surprise them if any changes appeared on the screen. For example, commands like "scroll up" or "swipe left" will navigate between pages; "play audio" will either play the recorded audio or activate text-to-speech; "like post" will add a post to their favorite screen. Creating a new post is also simple; users can select the photos from the photo library or take them from the camera, then users can use the audio aid to generate the post description, so they do not need to type any words. Audio playback is also available by pressing the play button, which allows users to be confident with the recorded messages and redo the recording if they encounter any unwanted voice (Fig. 5 (b)). All these features are built to increase older adult's comfortability of viewing and sharing their thoughts with their families or friends through our app.

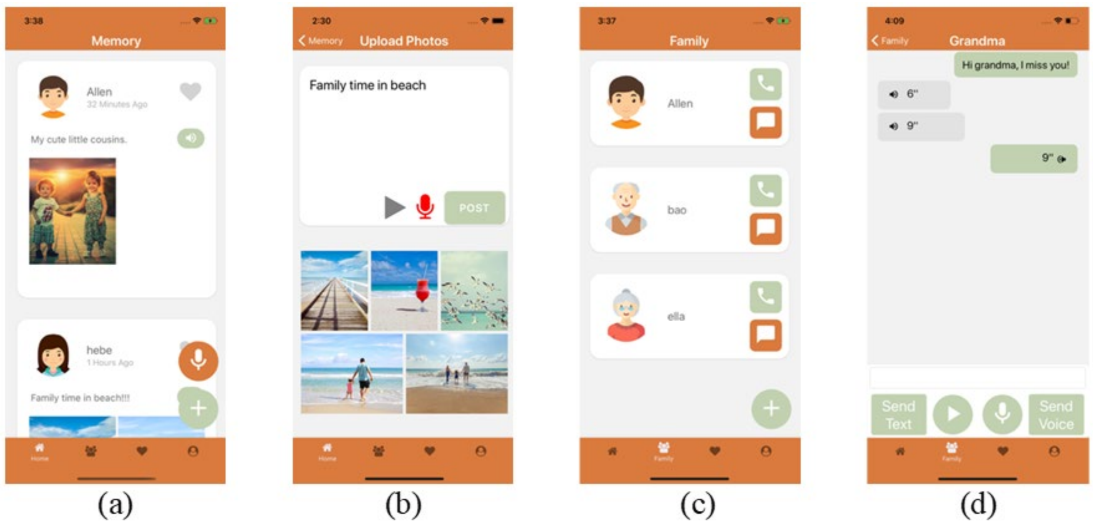

Fig. 5. (a) Home Screen; (b) Create Post Screen; (c) Family Contact Screen; (d) Chat Screen.

React-native voice recognition library is used to recognize the users' voice. The customized word extraction method detects the command words for every sentence the user spoke, where punctuation characters separate the sentences. Once a command word is detected, the app will temporarily stop receiving the voice input until the voice feedback and action are performed.

The contact screen in Fig. 5 (c) contains all the contact profile images that allow older adults to identify their contacts easily. Users can view a person's posts by pressing the person's contact card and making a phone call or video call with that person by pressing the phone button. The message button will lead to a private chat screen (Fig. 5 (d)) with the selected person. An audio button is available on the chat screen to free older adults from typing. Users can send either recorded audio or voice recognized text, in which users can then edit the recognized text if they encounter any mistakes in voice recognition. Besides that, when users click on any text message, the text-to-speech feature is available to play for the selected message. This feature serves as the visual aid for older adults who have difficulty in reading the text.

Evaluation

Due to COVID-19, we don't have the opportunity to visit senior centers in person and have our app tested by older users to receive feedback even though our IRB approval is ready. In the meantime, we did manage to have the app tested informally by a few older friends at home. Overall, they gave positive feedback on the senior-friendly interface. They also found the senior manager feature extremely useful as this functionality doesn't exist in common social media applications, and they felt comfortable letting someone they trust set up their account. But they did not find the voice navigation feature very useful as they've already had some experiences in using mobile applications and have no problem with performing simple gestures of swiping and pressing on the screen. With this feedback, we added new features to allow users to terminate the voice navigation through voice commands or simply turn off by press the voice button. Users can also keep the voice navigation on while using the screen interaction. We also received feedback suggesting new features, such as adding group chat and automatically cleaning chat messages after a selected time period. We aim to implement these functions to make our app comparable with the existing social media apps.

However, these feedbacks were from an informal study with a few users, therefore it does not hold statistical significance. We plan to conduct a formal study when the pandemic is over on having older adults and their family members test out and evaluate our application. We will first interview and conduct a pre-test survey with each older participant to understand their familiarity and former experiences with mobile applications and put them into three groups based on their technology familiarity. We plan to evaluate the usability of our app, particularly the voice navigation and senior manager feature. To evaluate the voice navigation efficiency and our word extraction method, we will analyze and compare the time an older user is required to navigate between posts through screen touching and voice navigation. We will also have the older participants try out the core functionalities and evaluate the difficulty of performing the tasks, e.g., locating and viewing a user's posts, messaging, and calling a user, and selecting a user as their manager. Furthermore, we will let the family members try out the senior manager interface to help their older family members set up contacts and reminders. At the end of the experiment, each participant will take the user experience survey to compare the experience on our app with the common communication methods or apps that they actively use. We will analyze all the collected feedback to improve our app design and features in the future.

Conclusion

The Share&Care app aims to increase older adults' comfortability with mobile communication and promote distanced family interaction. The UI and UX designs (Yun, 2016) of the app are tailored to older adults to increase their comfort in using our app and provide incentives for older adults who do not use mobile apps due to unfamiliarity or discomfort with technology to engage with our app. In particular, voice navigation and senior manager features were designed and implemented to ease older users' usage. By providing voice navigation with simple voice commands, we attempt to solve the problems that older users have in using touchscreen efficiently due to decreased hand dexterity. By providing the senior manager feature, which lets the older users designate someone they trust to manage their account, we attempt to save older users from complex operations on the app and closely connect them with their families. Furthermore, the managers can set up reminders for them remotely, either send one-time or periodic notifications to their phones to help them remember important dates or tasks.

In the future, we would like to further expand the senior-friendly features, by making our app totally hands-free and incorporating more voice feedbacks to make the older adults feel more confident in using our application; but this will be decided after the formal study if the voice interface is the best choice. As it could be difficult for older adults to find our application from the app store, we plan to compress our app into QR code and allow users to scan for downloading. We would also expand our multi-language feature to make it available for more languages.

Acknowledgments

This work is supported by a CEN Course Innovation Grant from Moxie Foundation. The first author and the last author are also supported by ODNI via the Intelligence Community Center for Academic Excellence (IC CAE) at Rutgers University (#HHM402-19-1-0003 and #HHM402-18-1-0007). An early version of this work has been presented at the 2020 IEEE MIT Undergraduate Technology Research Conference, October 9-11, 2020 (Chen et al., 2020).

Works Cited

AAL Programme, "AAL Home 2020.", www.aal-europe.eu/. Accessed 23 November 2020.

Banks, Marian R., and William A. Banks. "The effects of animal-assisted therapy on loneliness in an elderly population in long-term care facilities." *The journals of gerontology series A: biological sciences and medical sciences* 57.7 (2002): M428-M432.

Cacioppo, Stephanie, et al. "Toward a neurology of loneliness." *Psychological bulletin* 140.6 (2014): 1464-1504.

Cacioppo, John T., and Stephanie Cacioppo. "Older adults reporting social isolation or loneliness show poorer cognitive function 4 years later." *Evidence-based nursing* 17.2 (2014): 59-60.

Chen, Jin, Yu Ting Yang, Xiaohong Zhu, and Zhigang Zhu, "Share&Care: a senior-friendly family interaction application." 2020 IEEE MIT Undergraduate Technology Research Conference, October 9-11, 2020.

"Demographics of Mobile Device Ownership and Adoption in the United States." *Pew Research Center: Internet, Science & Tech*, Pew Research Center, 12 June 2019. www.pewresearch.org/internet/fact-sheet/mobile/. Accessed 5 August 2020.

Gao, Jing, and Andy Koronios. "Mobile application development for senior citizens." *PACIS*. 2010.

Jennifer Yeh, Shu-Chuan, and Sing Kai Lo. "Living alone, social support, and feeling lonely among the elderly." *Social Behavior and Personality: an international journal* 32.2 (2004): 129-138.

Kurniawan, Sri. "Older people and mobile phones: A multi-method investigation." *International Journal of Human-Computer Studies* 66.12 (2008): 889-901.

Marengoni, Alessandra, et al. "Prevalence of chronic diseases and multimorbidity among the elderly population in Sweden." *American journal of public health* 98.7 (2008): 1198-1200.

Pavlakis, Panagiotis, et al. "Intelligent mobile multimedia application for the support of the elderly." *2012 Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*. IEEE, 2012.

Perrin, Andrew. "Social media usage." *Pew research center* (2015): 52-68

Roberts, Andrew, et al. "The Population 65 Years and Older in the United States: 2016." *Census Bureau*, October 2018. www.census.gov/content/dam/Census/library/publications/2018/acs/ACS-38.pdf. Accessed 8 August 2020.

Sukmana, Septian Enggar, et al. "The Development of Berbakti: Elder Caring Mobile Application in Indonesia." *Journal of Applied Informatics and Computing* 3.2 (2019): 68-72.

Taylor, Harry, et al. "Loneliness in senior housing communities." *Journal of gerontological social work* 61.6 (2018): 623-639.

West, Loraine, et al. "65+ in the United States: 2010." *Census Bureau*, June 2014. www.census.gov/content/dam/Census/library/publications/2014/demo/p23-212.pdf. Accessed 8 August 2020.

Yun, You-Dong, Chanhee Lee, and Heui-Seok Lim. "Designing an intelligent UI/UX system based on the cognitive response for smart senior." *2016 2nd International Conference on Science in Information Technology (ICSITech)*. IEEE, 2016.

Developing a User-Centered Accessible Virtual Reality Video Environment for Severe Visual Disabilities

Kurtis Weir, Vinita Nahar, Amar Aggoun, Andrew Pollard
University of Wolverhampton, UK.
kurtisPW@hotmail.co.uk, vinita.nahar@wlv.ac.uk, a.aggoun@wlv.ac.uk, a.pollard@wlv.ac.uk
Fernando Loizides
Cardiff University, UK.
loizidesf@cardiff.ac.uk

Abstract

We address a timely issue of accessibility for visual information through the means of videos. Using emerging technologies (Head Mounted Virtual Reality Displays) and a user-centred design approach, we provide people with severe visual disabilities with a bespoke platform for accessing and viewing videos. We report on newly created test methods for measuring acuity within virtual spaces and reactions of impaired individuals, which informed our platform's design, to inform similar designs and allow testing and refinement for ecological and external validity. A prototype software for accessible virtual reality video viewing is presented, with a subsequent user evaluation to test the software, and a newer virtual reality head mounted display to determine usability while measuring how visually impaired users utilize elements in a virtual environment. We give guidance, based on empirical evidence, and advocate that although VR technologies are currently primarily targeted at a generic audience (gaming and entertainment), they can and should be further developed as assistive tools that enable independent living and increase the quality of life for those with disabilities, and specifically severe visual impairments.

Keywords

Virtual Reality, VR, Visual Disabilities, Video, Viewing, Accessibility

Introduction

Virtual Reality (VR) has experienced a resurgence, aided by its improving affordability, increased range of software, and hardware capabilities now matching user experiences needed for adoption. Head mounted displays (HMDs) are prevalent over similar VR technologies (Antycip; Horan et al.) due to their portability, immersion, and price (Oculus).

Mainstream focus of commercial VR headsets is entertainment, and although some work towards areas such as training (Barad; Elliman et al.) or therapy (Psious) exists, the technology is still at the research and development stage in regards to accessibility, missing mainstream adoption into consumer accessibility communities. HMDs are now advanced enough to be used as accessibility tools, enabling individuals with visual disabilities to live more independent lives through experiences otherwise restricted to them (Lee). An activity that our visually impaired participants desire greatly is the capability to watch TV again, being able to view videos comfortably.

Despite modern improvements, studies show that video player accessibility is still problematic for the elderly, who are the most likely to suffer from visual impairments (Villena et al.). Although there are existing ways to enhance video for impairments (Fullerton and Peli), the devices themselves (e.g. monitors, mobile) limit feasibility and effectiveness, with many impairments still restricting access to video.

The work presented in this paper, focuses on:

1) Creating a methodology and "benchmark" for assessing visual acuity within HMDs for visually disabled people with regards to viewing video content.
2) Presenting and assessing the usability of a VR HMD based platform for viewing videos, targeted at our visually disabled users with a follow up user evaluation.

Related Works

There are various types of HMDs today, most either video see-through (VST), or optical see-through (OST). VST devices (Massof and Rickman; Massof et al.) typically function by displaying a video feed to the user's eyes via an HMD, while OST devices (Microsoft; Google) overlay digital visuals over natural sight. Our work focuses on VST devices as they allow greater control over the visual feed via camera mounts, not relying as heavily on natural vision. Despite VST improvements, they still share many shortcomings of past devices (Harper et al.), and motion sickness is still problematic (LaViola Jr).

Research has been conducted combining VR and AR through reading live video feeds via camera mounts placed on VR devices (Zhao et al. 2015, 2016). Adaptability between multiple visual enhancements was praised by participants, and increased performances when compared to typical assistive tools, highlighting their suitability as visual aid tools.

There have been some attempts (Deahl; Relumino; GiveVision) at product releases for visual aiding headwear integrated with sight enhancing technology (eSight), although expensive costs and limited availability stagnate their adoption and accessibility to the public, a concern shared previously by Wolffsohn and Peterson, and a contrast to VR headsets today which have lowered their prices. Rather than expensive specialist devices, a cheaper VR headset could be purchased with accessibility software setup to assist with tasks. Reviews have shown that electronic vision aid devices are preferred and more effective than traditional optical devices, and that older CCTV attempts have shown increased performance in reading, yet there are too few studies to form strong conclusions (Jutai et al.; Moshtael et al.; Harper et al.). Looking at the conclusions of older research (Everingham et al. 1998, 2003; Peli et al. 1991, 1994) combined with VR today (Hwang et al.) could be the next step in accessibility for the visually impaired.

Discussion

Participant Selection & Test Environments

Prior to development of our specialist software, a preliminary test was conducted to inform us on how visually impaired users would observe and perform within a virtual environment. Our study selected 9 participants that were classed as “Low Vision” (Corn and Lusk) and were all supported in their day to day lives due to the level of their visual disability in various ways. The type of condition did not matter if each participant fell into this LV category. Generally, these participants were older adults either middle aged or older, as is the typical age bracket with most severe visual impairments.

To understand the different visual elements of playing video within a 3D space, and what might affect these, a set of measure-able variables were defined that could be fed into the development of our software. Following discussions with an optometrist, we devised 10 test variables that would allow us to cover different visual aspects within a 3D space. These were Color, Location, Speed, Contrasting Movement, Distance/Depth, Detail, Contrast, Size, Brightness, and Distinction. As the way VR devices display imagery is unique and differs from both natural vision and standard 2D screens, utilizing an existing test could yield inequitable results, prompting us to design our own methods to test these variables.

These developed tests allowed us to observe how LV users might utilize an HMD and its environments, what is beneficial to them, and what limitations are present and could be addressed in further developments. This feedback fed directly into the development of the subsequent video player as design considerations were made based on our participant's reactions. Tests presented here are designed to cover one or more of our original variables listed above.

Object Tracking

This first test was designed to examine the user's ability to identify and track a simplistic object within VR. The user follows a green sphere while facing directly forward, testing their perception at differing horizontal and vertical positions, as well as distance as the sphere travels further away and back again (see fig. 1). This allows us to determine any damaged visual cones.

Results showed that 5 users with damaged visual cones could only see portions of the sphere and that vertical tracking was easier than horizontal. Background colors blended with the sphere, but its color could still be determined and was easiest next to white. All but 1 user could fully track the sphere to the end, with 1 losing vision beyond 15m ranges temporarily.

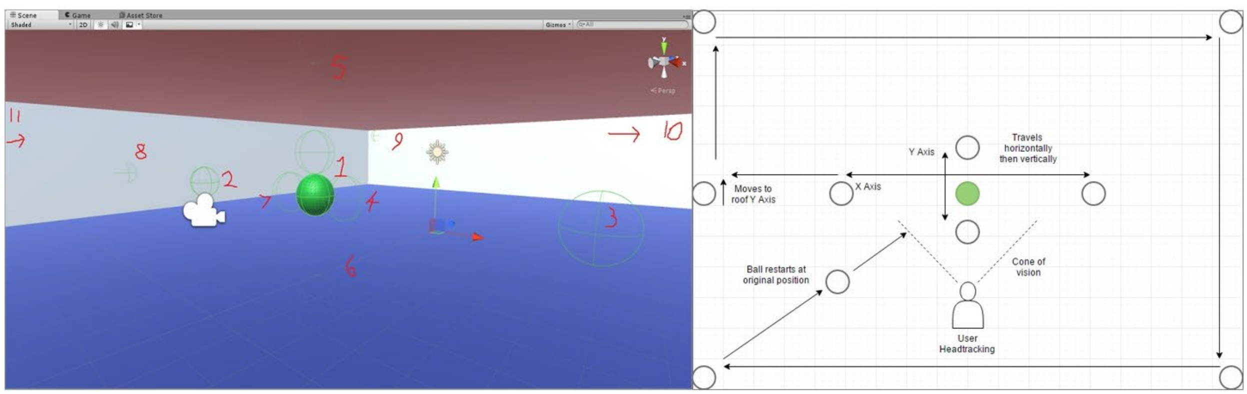

Fig. 1. The left image shows the Unity scene of the Object tracking test, and the right image the wireframe diagram of the sphere's movement pattern.

Movement

This test was designed to gauge the perception of users differentiating between multiple visual elements with varying levels of movement, simulating visual noise. Multiple objects move in front of the user while they attempt to identify them, count them, describe their movement, and list objects in order of speed (see fig. 2).

Results showed that 2 users struggled with identifying colors, 1 struggled with counting all objects due to limited vision cones, and 1 struggled with distinguishing between varying speeds. Remaining users accurately identified between fastest and slowest objects. Those that struggled were distracted by visual noise, with 1 seeing static objects as moving.

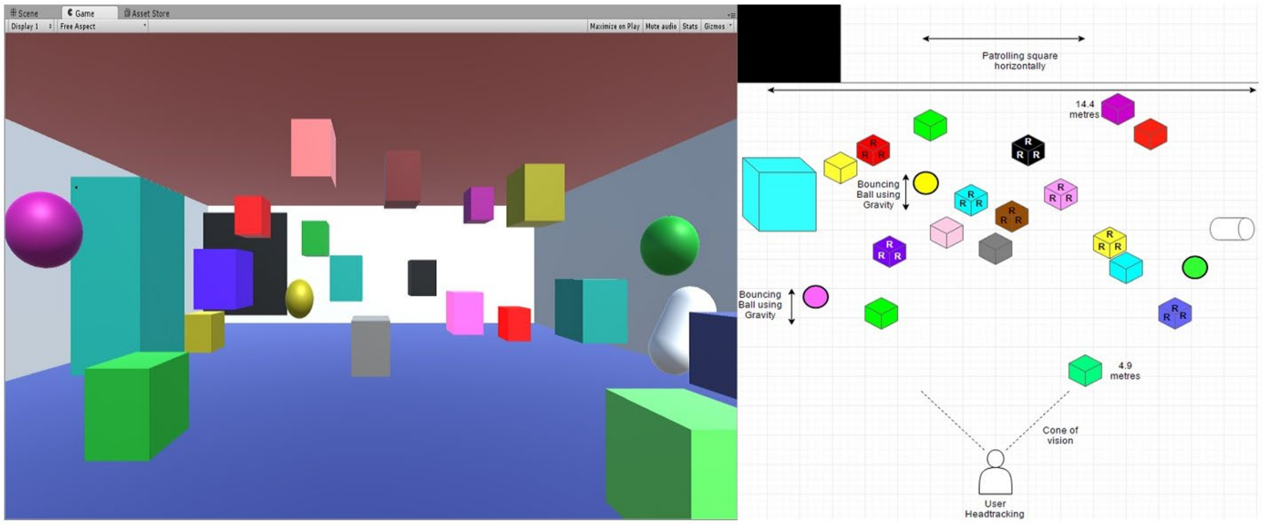

Fig. 2. The left image shows the Unity scene of the Movement test, and the right image the wireframe diagram of each object's characteristics.

Gaze Ability

This test was designed to observe cognitive ability within VR with simple task solving, and the ability of users suffering from central vision loss, such as AMD (Age-related macular degeneration). Red arrows direct the user to subsequent arrows changing green once a central hit scan is detected, until an entire sequence has been followed (see fig. 3). Collision can be calibrated to compensate to the degree of central vision missing.

Results showed that all users but 1 could complete the full test's sequence. This user struggled due to central vision loss, particularly identifying arrows placed higher up. 2 users required central vision calibration to complete the test, benefitting them afterwards. 1 user observed the arrows as moving, despite being static.

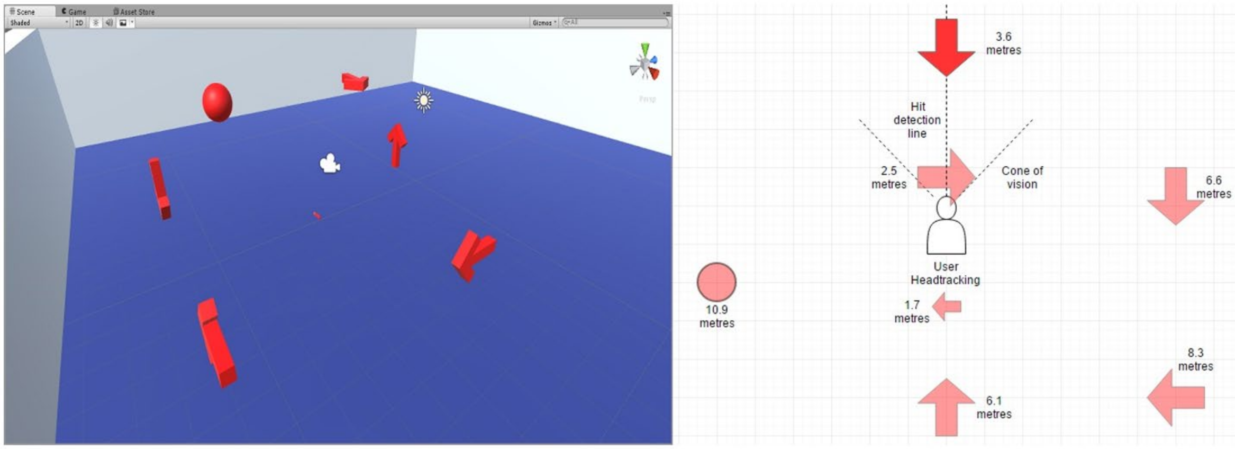

Fig. 3. The left image shows the Unity scene of the Gaze test, and the right image the wireframe diagram of each arrow's location.

Distance Color

This test was designed to examine the user's ability to detect and correctly identify color over varying distances. Four green 3.3753 m cubes of different colors were presented at 3.9m from the user, and after each attempt of identifying each color these cubes move a further 2m distance, changing color until every sequence is completed (see fig. 4).

Results showed that 2 users struggled with colors, with 1 unable to see color and beyond 17.5m from one eye. One user's perception was affected by which side of their head a cube was at. As expected, distance affected the accuracy of color detection, particularly those with central vision loss, highlighting the need for control over distances.

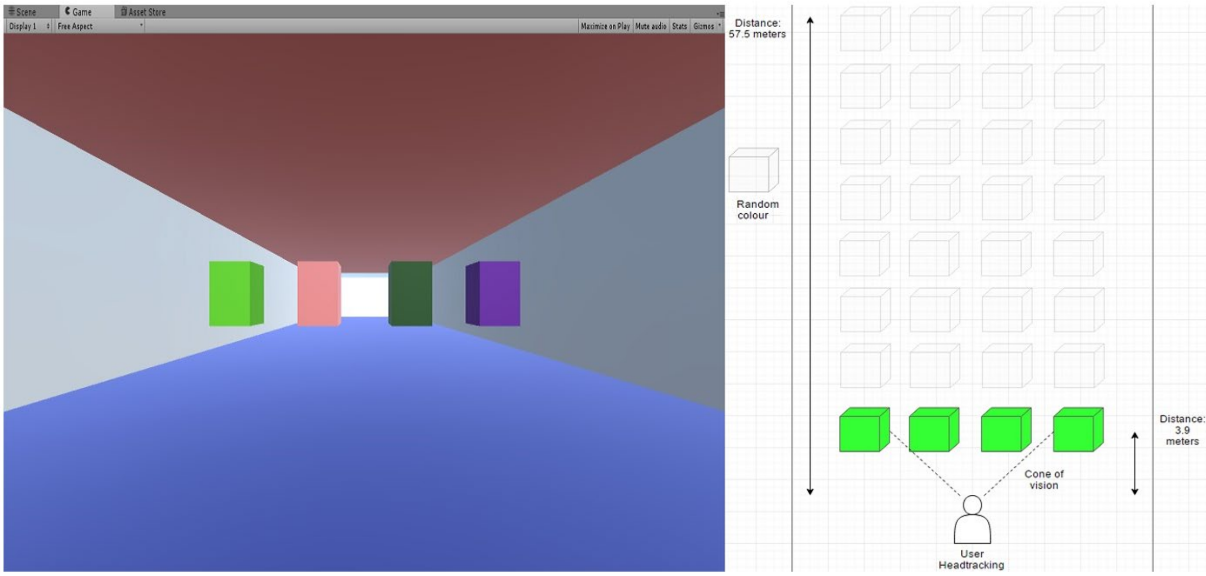

Fig. 4. The left image shows the Unity scene of the Distance test, and the right image the wireframe diagram of the scene.

Size Detection

This test was designed to gauge the user's ability to identify dimensions in nearly identical shapes. 4x5 rows of red cuboids were displayed, some perfect cubes, some distorted by an axis or entirely scaled, with the user asked to identify and spot out irregular cubes and describe any differences they could perceive (see fig. 5).

Results showed that no user could fully complete the test sequence, and 1 user could not detect any size differences. One user described cuboids as wobbly and blurry and could not see differences until highlighted to them. Exaggerated objects were best detected, and VR development should ensure elements are contrasting and clearly defined for accessibility.

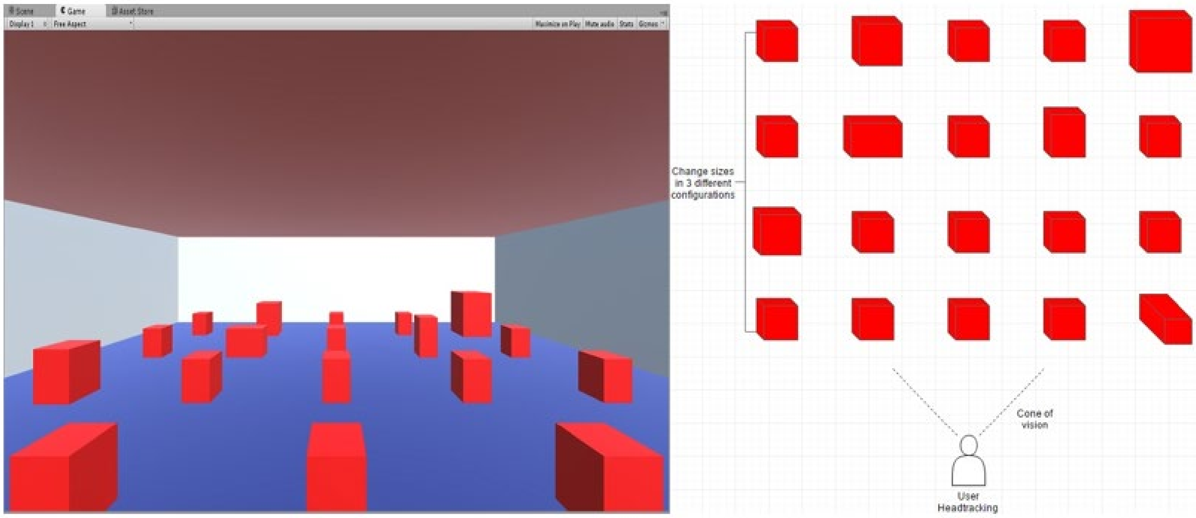

Fig. 5. The left image shows the Unity scene of the Size test, and the right image the wireframe diagram of each cube's size.

Depth Perception

A final test was designed to gauge depth of perception as a participant reported having stereo-blindness. The first half of the test has the user examine a 3D and 2D circle and determine the difference in a 3D space, while the second half has the user look at several pillars distances behind each other at longer distances while describing their perception of depth (see fig. 6).

Results showed that 6 users could correctly perceive depth, 2 were unsure and struggled describing, and 1 could not at all. One user commented that their depth has been long gone, and they were able to correctly observe it again through the headset, but we could not replicate these same results with any other participant.

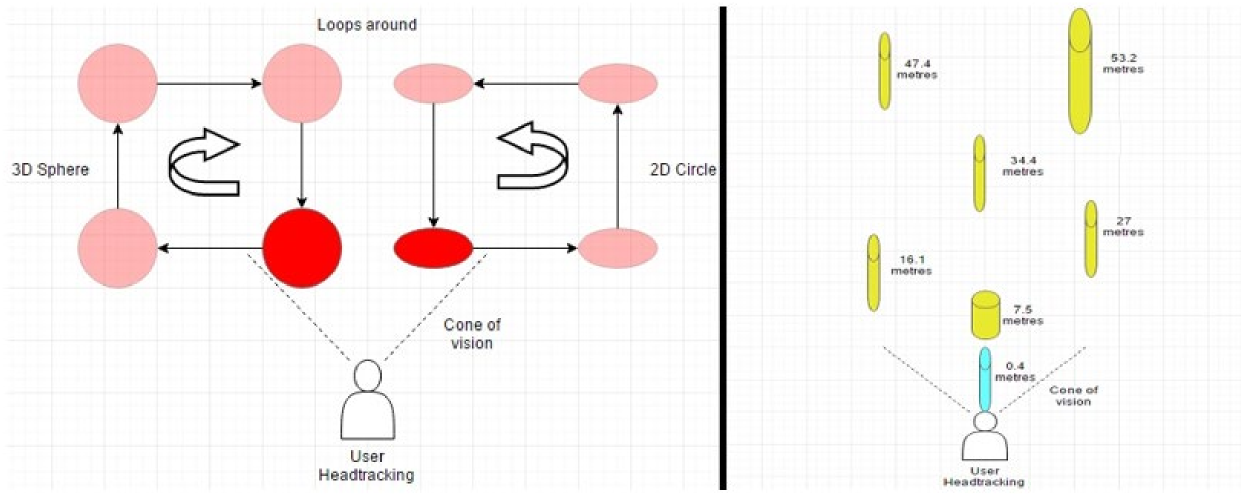

Fig. 6. The left image shows the wireframe diagram of the first part of the Depth test, and the right image the wireframe diagram of the second part.

Summary

Findings showed that color accuracy was manipulated by both visual noise and distance, vertical tracking was easier for tracking an object, but worse for finding an object for AMD sufferers, color accuracy degraded near 17m distances, size differences between near identical objects were difficult to perceive, and that depth perception might be affected within VR but limited results were produced. It is worth noting that the tested headset produces clearer visuals when looking at the center of the lenses, which is a limitation. Reactions from participants were very positive, with most suggesting they could not perform these tasks without the headset. This gives us a framework for what impaired users are capable of within a VR headset, informing future developments towards VR accessibility, such as our software.

Software Prototype & User Evaluation

Following the preliminary test, a VR video player was developed based on participant results and feedback using the Unity engine and tested via a Pimax 5K Plus VR headset (Pimax). Our findings highlighted that control over distance and placement (such as elevation to compensate for vision cones) would improve the viewing capabilities of visually impaired users. Contrasting colors, visual noise, and distinguished elements are also key and can be emphasized within a digital environment, showing the benefits of a VR system. The application allows users to select and setup their own 3D virtual living environment and view videos that can be manipulated with several enhancements and size adjustments. Using motion controllers, users watched videos within the 3D environment which could be customized for both preference and accessibility, such as enabling virtual lights from specific directions or torches, or to place and re-position both the video and the environment. 2D, 3D, and 360-degree video formats are supported, and can be embedded from external websites. The user can resize the video player by stretching it (see fig. 3 (left)). The video player can be moved around freely and frozen virtually grabbing and releasing it via motion controllers. Curvature of the screen can be adjusted to assist with viewing angles. The light and color levels of the video can be adjusted for accessibility needs. Overall brightness and contrast of the video, or of the entire environment, can be adjusted to assist with clarity and to aid specific conditions such as photosensitivity. Specific color adjustments can be made to help with conditions such as color-blindness (see fig. 7 (right)). Most options are also available through voice commands spoken via an embedded microphone into the VR headset, thus requiring less dexterity from the user.

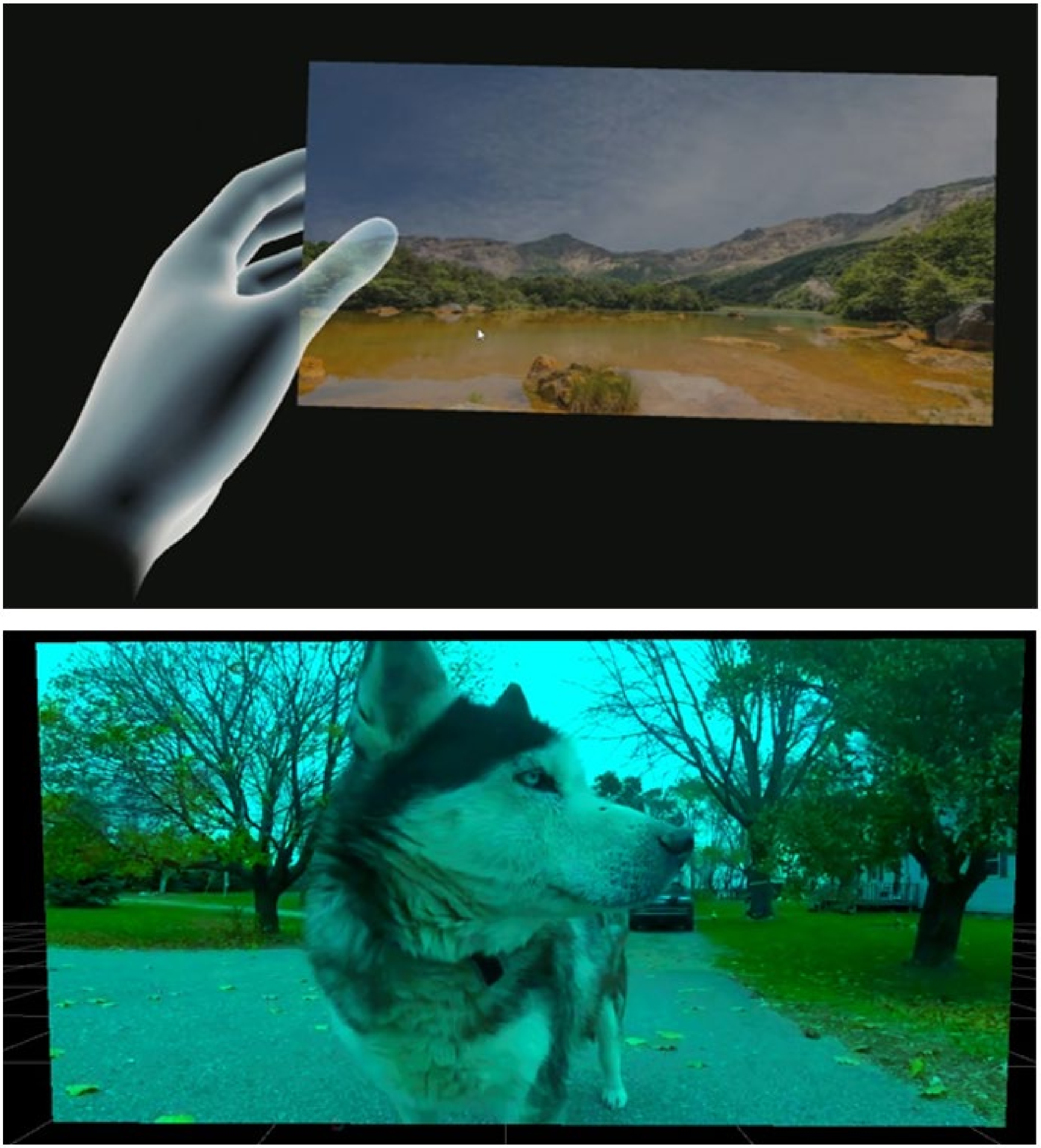

Fig. 7. (Top) A video shrunk to the user's hand in our prototype; (Bottom) A video with colors adjusted to compensate for Protanopia color-blindness.

The environment can be manipulated using the motion controllers to grab, re-arrange, and place objects. Configurations are saved and loaded, and the user can choose how complex or simple the environment is to increase immersion factors, or to reduce visual noise (see fig. 8).

Fig. 8. An example bedroom environment configuration.

After the prototype software was developed, a second study was conducted trialing a further 11 participants still classed as Low Vision. We recorded configurations each participant selected while using the video player, notably their **contrast**, **brightness**, **distance**, **video size**, and **rotation settings**, asking for their preferred preferences. A follow up interview and questionnaire was conducted for feedback.

Results showed that 7 users preferred an increased brightness, while 4 were sensitive to light with 3 of these preferring brightness unchanged, and 1 the brightness reduced. Seven preferred an increased contrast, with six of these by a significant amount and one only minor. Three preferred no changes, and one reduced contrast levels. Average distance for preferred viewing was 0.63m, with 3 preferring 0.3-0.4m ranges and 3 preferring 0.78-0.84m ranges. Ten users kept the video size at its default 60x30cm at 0.5m distance, while one increased this by 20%. Similarly, users did not make significant changes to the video rotation, while 1 rotated it 20 degrees upwards and placed it 120 degrees forward to the right of their vision.

In summary, participants displayed an affinity towards the ability to adjust the brightness and contrast dynamically themselves. Participants expressed the biggest improvement to visual clarity in manipulating brightness, and to a lesser extent the contrast. The preferred location for the video player was left close to its default position (0.5m away from eye level).

Observational and verbal feedback from our test group showed that natural real-world interactions via motion and voice controls were understandable and preferred by impaired users. Participants quickly understood and enjoyed interactions by gripping triggers via HTC Vive Wands (HTC) to move objects. Participants expressed they would utilize an accessibility headset daily if it were lighter and more comfortable to wear.

Conclusions

In this paper we successfully developed user-informed accessibility work in VR to enable persons with disabilities to watch videos, or to improve their video watching experience, as part of increasing their independent living capabilities. The presented tests give us a greater clarity of what visually impaired users are capable of within VR environments, and their level of perception amongst tested variables. This type of study is important, as current top commercial VR headsets are not designed with the visually impaired in mind, yet the technology could greatly enhance their lives. These findings were then used as the foundation of our bespoke highly adaptable VR video viewer. This video viewer designed for accessibility is the first of its kind and aims to highlight and promote the potential for using VR technology as accessibility tools.

Feedback from our user evaluation highlighted that the ability to re-position virtual objects dynamically via motion controls was very well received by impaired users, as details such as modifying viewing angles to combat damaged visual cones effortlessly allowed for easier viewing. Our evaluation validated the usefulness and usability of such a device combined with specialist software, as well as where improvements could be made. Our participants gave us a baseline for what settings persons with severe visual impairments would use to view and operate a video player within a VR environment, and what aspects that affect their acuity. Our research showed that video elements were perceived well with persons of severe visual impairments, and flagged some concerns with field of view limitations, (particularly those with AMD) and brightness settings being a concern.

In future works we aim to develop the software further by integrating live video with enhancements via camera mounts and compare performance to natural vision.

Works Cited

Antycip. “VIRTUAL REALITY — VR CAVE.” https://www.antycipsimulation.com/vr-cave/, 2020. [Online; accessed 4-April-2020].

Barad, Justin. “Virtual and augmented reality can save lives by improving surgeons’ training.” https://www.statnews.com/2019/08/16/virtual-reality-improve-surgeon-training/, 2019. [Online; accessed 21-April-2020].

Corn, AL and Lusk, KE. “Perspectives on low vision.” 2010.

Deahl, D. “Samsung launches Gear VR app Rel´um˘ino that aids the visually impaired. The Verge.” https://www.theverge.com/tech/2017/8/21/16178936/samsung-gear-vr-oculus-relumino-low-vision-free, 2018. [Online; accessed 16-January-2020].

Elliman, James, Loizou, Michael, and Loizides, Fernando. “Virtual reality simulation training for student nurse education.” *2016 8th International Conference on Games and Virtual Worlds for Serious Applications (VS-Games).* IEEE, 2016, 1–2.

eSight. “eSight FAQ - Electronic Glasses for the Legally Blind.” https://www.esighteyewear.com/faq, 2018. [Online; accessed 16-January-2020].

Everingham, MR, Thomas, Barry T, and Troscianko, Tom. “Wearable mobility aid for low vision using scene classification in a Markov random field model framework.” *International Journal of Human-Computer Interaction* 15.2 (2003): 231–244.

Everingham, MR, Thomas, BT, and Troscianko, T. “Head-mounted mobility aid for low vision using scene classification techniques.” *International Journal of Virtual Reality* 3.4 (1998): 1–10.

Fullerton, Matthew and Peli, Eli. “Post-transmission digital video enhancement for people with visual impairments.” *Journal of the Society for Information Display* 14.1 (2006): 15–24.

GiveVision. “Give-vision.” https://www.givevision.net/, 2018. [Online; accessed 16-January-2020].

Google. “Discover Glass Enterprise Edition.” https://www.google.com/glass/start/, 2020. [Online; accessed 19-January-2020].

Harper, Robert, Culham, Louise, and Dickinson, Christine. “Head mounted video magnification devices for low vision rehabilitation: a comparison with existing technology.” *British journal of ophthalmology* 83.4 (1999): 495–500.

Horan, Ben, Sevedmahmoudian, Mehdi, Mortimer, Michael, Thirunavukkarasu, Gokul Sidarth, Smilevski, Stephen, and Stojcevski, Alex. “Feeling Your Way Around a CAVE-Like Reconfigurable VR System.” *2018 11th International Conference on Human System Interaction (HSI).* IEEE, 2018, 21–27.

HTC. “VIVE — Controller.” https://www.vive.com/uk/accessory/controller/, 2020. [Online; accessed 3-March-2020].

Hwang, Alex D and Peli, Eli. “An augmented-reality edge enhancement application for Google Glass.” *Optometry and vision science: official publication of the American Academy of Optometry* 91.8 (2014): 1021.

Jutai, Jeffrey W, Strong, J Graham, and Russell-Minda, Elizabeth. “Effectiveness of assistive technologies for low vision rehabilitation: A systematic review.” *Journal of Visual Impairment & Blindness* 103.4 (2009): 210–222.

LaViola Jr, Joseph J. “A discussion of cybersickness in virtual environments.” *ACM Sigchi Bulletin* 32.1 (2000): 47–56.

Lee, A. "A rare disease robbed me of my sight. VR brought it back. Alphr." http://www.alphr.com/virtual-reality/1008932/vr-vision-loss-sight-blindness, 2018. [Online; accessed 13-January-2020].

Massof, Robert W, Baker, Frank H, Dagnelie, Gislin, DeRose, Joseph L, Alibhai, Suleiman, Deremeik, James T, and Ewart, Craig. "Low Vision Enhancement System: Improvements in Acuity and Contrast Sensitivity: 3:00 pm (LV-313)." *Optometry and Vision Science* 72.12 (1995): 20.

Massof, Robert W and Rickman, Douglas L. "Obstacles encountered in the development of the low vision enhancement system." *Optometry and Vision Science: Official Publication of the American Academy of Optometry* 69.1 (1992):32–41.

Microsoft. "Microsoft HoloLens — Mixed Reality Technology for Business." https://www.microsoft.com/en-CY/hololens, 2019. [Online; accessed 4-April-2020].

Moshtael, Howard, Aslam, Tariq, Underwood, Ian, and Dhillon, Baljean. "High tech aids low vision: a review of image processing for the visually impaired." *Translational vision science & technology* 4.4 (2015): 6–6.

Oculus. "All-In-One VR." https://www.oculus.com/quest/, 2019. [Online; accessed 7-April-2020].

Peli, Eli, Goldstein, Robert B, Young, George M, Trempe, Clement L, and Buzney, Sheldon M. "Image enhancement for the visually impaired. Simulations and experimental results." *Investigative ophthalmology & visual science* 32.8 (1991): 2337–2350.

Peli, Eli, Lee, Estella, Trempe, Clement L, and Buzney, Sheldon. "Image enhancement for the visually impaired: the effects of enhancement on face recognition." *JOSA A*11.7 (1994): 1929–1939.

Pimax. “Pimax 5k PLUS.” https://store.pimaxvr.com/shop/product/pimax-5k-plus-120hz-19, 2020. [Online; accessed 20-January-2020].

Psious. “Virtual Reality for mental health.” https://psious.com/, 2020. [Online; accessed 21-April-2020].

Relumino. “relumino — Timeline.” https://www.samsungrelumino.com/timeline-en, 2018. [Online; accessed 16-January-2020].

Villena, Johana M Rosas, Ramos, Bruno C, Fortes, Renata PM, and Goularte, Rudinei. “An accessible video player for older people: Issues from a user test.” *Procedia Computer Science* 27: 168–175.

Wolffsohn, James S and Peterson, Rachael C. “A review of current knowledge on Electronic Vision Enhancement Systems for the visually impaired.” *Ophthalmic and Physiological Optics* 23.1 (2003): 35–42.

Zhao, Yuhang, Szpiro, Sarit, and Azenkot, Shiri. “Foresee: A customizable head-mounted vision enhancement system for people with low vision.” *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility.* ACM, 2015, 239–249.

Zhao, Yuhang, Szpiro, Sarit, Knighten, Jonathan, and Azenkot, Shiri. “Cue-See: exploring visual cues for people with low vision to facilitate a visual search task.” *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing.* 2016, 73–84.

Usability of Audio Matrix Microtasks for Screen Reader Users

Ying Zhong[1], Masaki Matsubara[1], Makoto Kobayashi[2], Atsuyuki Morishima[1]

[1]University of Tsukuba, [2]Tsukuba University of Technology

ying.zhong.2018d@mlab.info, masaki@slis.tsukuba.ac.jp,

koba@cs.k.tsukuba-tech.ac.jp, morishima-office@ml.cc.tsukuba.ac.jp

Abstract

Crowdsourcing provides people with new work opportunities; it involves the concept of microtasks. Microtasks have been proven to be a promising work option for visually impaired people. Unfortunately, most real-world microtasks are designed for sighted people. The few studies on designing microtasks for visually impaired people focus only on transcription tasks; however, there can be additional types of microtasks that visually impaired people can perform using screen readers. Therefore, this study focuses on developing better microtask designs for visually impaired people. We focus on audio matrix tasks in which audio clips are placed in a tabular format. This study explains our approach, provides theoretical background for improving consistency in a given task, and presents the results of experiments conducted with visually impaired people. The results show better result quality and lower working time for visual impaired workers when performing the tasks.

Keywords

Blind/Low Vision, Web, ICT, Crowdsourcing

Introduction

Crowdsourcing is one of the technologies that provide people with new work opportunities. People without a traditional job can find crowdsourcing jobs and earn money. An important concept used in certain crowdsourcing is microtasks, which are tasks performed in a short period of time. However, most designs of real-world microtasks are focused on sighted people. There is existing literature claiming that microtasks are promising for visually impaired people to work (Vashistha et al.). Unfortunately, such work has focused only on transcription tasks; however, there may be additional types of microtasks that visually impaired people can perform using screen readers.

Therefore, this study attempts to further expand the opportunity for screen reader users to work as crowd workers by exploring a design space of more general microtasks with textual data. We first conducted a survey (Zhong et al.) with 91 visually impaired crowd workers on the usability of crowdsourcing platforms. In the survey, we found some difficulties faced by them. Visually impaired crowd workers can perform limited tasks in actual crowdsourcing platforms. Approximately 8% (8) of the visually impaired crowd workers, who use screen readers, have difficulty performing microtasks with visual content, such as text and images. Although they can listen to the text to complete tasks, the actual tasks on crowdsourcing platforms are not easy to understand. Furthermore, in the survey, we also found that time is an important factor for them. Therefore, we aim to develop a text arrangement for microtasks that improves the performance of visually impaired workers, who rely on hearing the instructions, in terms of time required and result quality.

We assume that task instructions and any data related to the task are given in an audio matrix, which is an N-dimensional space with coordinates, where an audio clip can exist at every location. Given an audio matrix microtask, workers can navigate the matrix physically, such as by pressing arrow keys; hear the audio matrix at each location; and submit task results.

The novelty and contributions of this study are as follows:

Case studies on audio matrix tasks for screen reader users: We devise consistency-conscious design of audio matrix microtasks for visually impaired workers and verify the performance of the design in terms of the result quality and time required.

Experiment with a large number of people: In addition to the case studies, we conduct large-scale experiments with people who emulate users of screen readers. We verify the performance of our consistency-conscious design with axis consistency for audio matrix microtasks.

Theoretical contribution: We showed that there is a theoretical framework based on the information capacity theory (Abu-hamdeh et al.), which defines the equivalence between data schemata, to systematically obtain audio matrix microtask formats with the axis consistency.

Discussion

Related Work and Comparison with Existing Services

Many people with disability work as crowd workers on crowdsourcing platforms (Zyskowski et al. and Calvo et al.). There are some studies that focus on tasks that are more accessible for visually impaired crowd workers. BSpeak is a crowdsourcing platform developed for visually impaired people (Vashistha et al.). Crowd workers can perform transcription tasks using voice-based implementation. Salisbury et al. recruited sighted crowd workers to simulate visually impaired workers for studying the manner in which crowdsourcing can be used both for evaluating the value provided by existing automated approaches and allowing workflows that provide scalable and useful alt text to visually impaired users.

To the best of our knowledge, our study is the first to focus on tasks that are in a tabular format (Jiang et al. and Wang et al.) for users of screen readers, which is a more general form for describing tasks. This research also aims to expand microtasks to visually impaired users. We recruit both sighted crowd workers to simulate visually impaired workers and visually impaired workers to test our microtask design.

Theoretical Formalization for Designing Audio-Matrix Tasks

Audio Matrix. An audio matrix task is a task in which information is represented in the form of an audio matrix. An audio matrix task assumes that workers have the means to navigate the locations of the N-dimensional space defined by the matrix. For example, users may be able to navigate using cursor keys, gestures, etc. In addition, there may be shortcut commands to go to particular locations, such as the home key to go to the root location (1, . . . , 1).

Definition 1 (Audio Matrix) *An N -dimensional audio matrix is represented by* $\{a_{l_i}\}$ *where* $l_i = (l_{i1}, \ldots, l_{in})$ *represents a location in the matrix and each* a_{l_i} *is an audio clip or null.*

Task Schema. Each task generator is associated with a *task schema s*, which represents the common logical format of the task the generator generates. In this paper, we use a tree representation for the task schema. Figure 1 represents an example of task schema.

Definition 2 (Task Schema) *A task schema s is a rooted tree* $s = (V, E, content)$ *that represents a logical structure of the task. Here,* $V = \{root\} \oplus Intermediates \oplus Leaves$, *and* $content : V - \{root\} \rightarrow String$ *is a function where* $content(v)$ *returns the explanation of the content of v, which may be preceded by * (repetition mark).*

The problem is dealt in the information capacity theory (Abu-hamdeh et al.), where the equivalence of different database schemas is discussed. Let T (s) denote the set of all valid instances (i.e., tasks) of schema s, where a valid instance of a schema conforms to the structure and constraints defined by the schema.

Definition 3 *An information capacity preserving mapping (or information preserving mapping) between the instances of two schemas* s_1 *and* s_2 *is a total, injective function* $f : T(s_1) \rightarrow T(s_2)$.

Definition 4 *If* $f : T(s_1) \rightarrow T(s_2)$ *is an information capacity preserving mapping, then* s_2 *dominates* s_1 *via f, denoted* $s_1 \leq s_2$.

For example, the task schema s_1 in Fig. 1 (Left) and s_2 in Fig. 1 (Right), $s1 \leq s_2$ holds. In the visual tasks in Fig. 2 top-left and bottom-left, we give a real example. We set the portrait is Dimension 1 and landscape is Dimension 2.

Consistency-Conscious Design of Audio Matrix Tasks

Naive mapping of the information of visible microtasks to an audio matrix may be confusing. Given the results of our preliminary attempts, we hypothesized the following:

Hypothesis [Axis Consistency]. Inconsistency of the semantics in navigation of each dimension leads to confusion, which in turn leads to lower quality task results.

For example, the task shown on the left of Fig. 1 violates the consistency constraint; here, the significance of values in Dimension 1 changes with the location. At certain instances, these values refer to the attribute of different objects and at other, they correspond to task-level instructions (questions and operations). We label each node in the schema with “t-item” if it is a task-format-level data item or with “o-item” if it is an object-level data item.

The hypothesis is consistent with a well-known result in cognitive science: a balanced configuration exists if the attitudes towards the parts of a causal unit are similar.

Generating Consistency-Conscious Design. First, we use a systematic translation of equivalent schema (task formats in our context) and obtain a set of audio matrix tasks that are equivalent to each other. Then, we choose a set of understandable audio matrix tasks, where the semantics of navigation for every axis is consistent. This process can be automated if we present the task structure in a tree form as in (Abu-hamdeh et al.) and annotate every component with a label that explains its meaning, such as t-item and o-item, or more specifically, the hotels to be compared, their metadata, etc. Then, for each equivalent task represented in a tree form, we check whether the labels appearing in the same dimension are consistent.

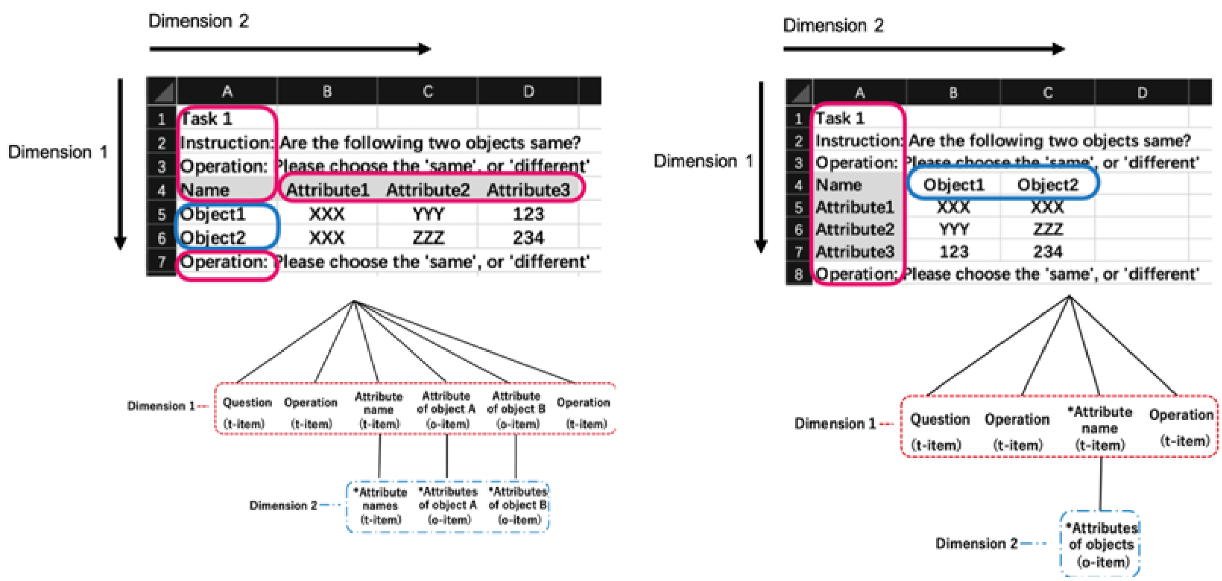

Fig. 1. Data Structure Used in Tasks.

Vertical and horizontal axes of the data table represent Dimensions 1 and 2, respectively. Here, t-item is the description of the tasks and o-item is the data item to be compared. (Left) Data structure of the original (inconsistent) task shown in Fig. 3 and Fig. 4. (Right) Data structure of task with axis consistency shown in Fig. 3 and Fig. 4. The data structure of task with axis consistency is simpler than that of original task. Based on information capacity theory, our principled framework can generate a task with axis consistency from the original data structure.

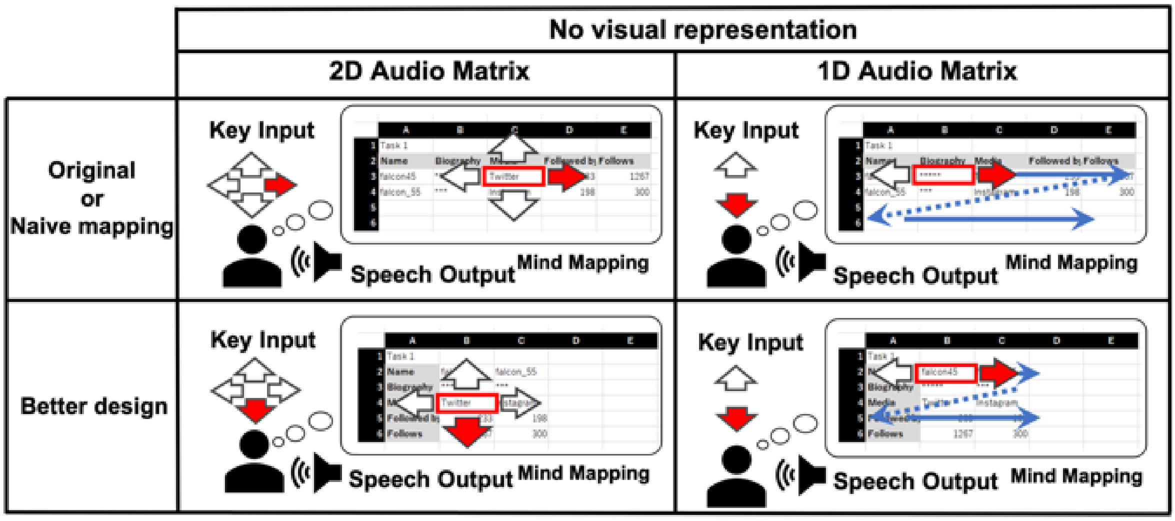

Fig. 2. Example of Four Task Representations Used in Our Experiment.

(Top-left) Naive mapping 2D audio matrix. (Top-right) Naive mapping 1D audio matrix. (Bottom-left) Axis consistency 2D audio matrix. (Bottom-right) Axis consistency 1D audio matrix. In these 2D audio matrix tasks, the user can navigate locations in vertical and horizontal directions with up/down and left/right arrow keys, respectively.

Experiments and Results

Participants

To investigate usability of audio matrix microtasks, we conducted case studies with three visually impaired workers who use a screen reader. In addition, we conducted experiments with 180 sighted people who emulated screen reader users. These experiments were approved by IRB committee of University of Tsukuba. All sighted participants were recruited via Amazon Mechanical Turk. The visually impaired workers were recruited directly. In the experiment, four task representations were designed and compared with regards to two factors: 2D vs. 1D and consistent vs. inconsistent (Fig. 2).

2D vs. 1D audio matrix design: From our previous interview, it was clear that visually impaired people are good at using Microsoft Excel. Therefore, we designed 2D audio matrix tasks on an application similar to Excel. In addition, visually impaired people usually use swipe action to operate smartphones; they acquire 2D information as 1D sequential audio information (Watanabe et al. and Sanderson et al.). Assuming using a smartphone is appropriate for microtasks, based on these facts, we designed 1D audio matrix tasks with less operation which transfer from 2d tasks (Audio Matrix Task).

Consistency-conscious design: We used the consistency-conscious design of audio matrix microtasks for different types of content in different dimensions. The navigation of tabular data was found to be easier. Our experimental design is illustrated in Fig. 2, where an audio matrix task is conceptually represented in a worker's mind in a table-like format. For example, in the bottom-left task in Fig. 2, when a worker presses the same key, such as the right arrow key on the keyboard, if it always implies that the worker is required to compare the values of the same attribute of different objects, the task will be easier to understand. However, the task in the top-left in Fig. 2 does not guarantee that consistency. A worker needs to press the down arrow key to not only compare the values but also read the task instruction.

Tasks

Data. We focused on tasks that are associated with tabular data, such as entity resolution tasks, which appear in many real-world applications and are widely seen on online crowdsourcing platforms. The proposed method assumes that the tasks are associated with tabular data as often required by entity resolution tasks. We attempt to find a better design for this type of task.

Entity resolution tasks that use human-powered joint operation have received considerable attention (Christen). There are three different types of data structures that use human-powered joint operation (Mudgal et al.): (1) Structured: Data are structured (e.g., tabular data). (2) Textual: Data are not structured (includes text entries). (3) Dirty: Data are structured but attribute values may contain missing values. In our experiment, we used structured and dirty data. We chose dirty data because it contains missing values and is more difficult to judge.

Procedures. Participants were asked to perform two types of tasks: user matching and product comparison tasks. Fig. 3 shows the examples of user matching tasks. We asked crowd workers to compare two social media accounts with attributes, few of which were missing, to decide whether they belonged the same account or not. Fig. 4 shows the examples of product comparison tasks. We asked crowd workers to choose the most cost-effective product.

Results

The important findings of this study are as follows. First, for screen reader users, the performance was improved using a consistency-conscious design of audio matrix microtasks. We found the time required for axis consistency 2D audio matrix tasks were shorter; the participants preferred axis consistency 2D audio matrix tasks based on the interviews (See Table 1 and Table 2). Second, we empirically confirmed that axis consistency is an essential factor for microtask design without visual representations. The accuracy rate of sighted (simulating visually impaired) crowd workers also improved significantly (See Fig. 5 and Table 3).

Conclusions

This study addressed the problem of developing better microtask designs for visually impaired people, provided a theoretical background, and also presented the results of our experiments; the results show better result quality and lower time required for visual impaired workers to perform the tasks. In future work, we will explore additional datasets or languages and automate the task.

Acknowledgment

This work was partially supported by JST CREST Grant Number JPMJCR16E3 and JSPS KAKENHI Grant Number 17K20022, Japan.

Task 1					
Instruction: Are the following two the same social media users?					
Operation: Please choose 'the same' or 'different'.					
User name	Bio	Full name	Number of posts	Followed by	Follows
John	The world is beautiful.	John Smith	50	500	30
Mary		Mary Wang	20	200	10
Operation: Please choose 'the same' or 'different'.					

Task 1			
Instruction: Are the following two the same social media users?			
Operation: Please choose 'the same' or 'different'.			
User name	John	Mary	
Bio	The world is beautiful.		
Full name	John Smith	Mary Wang	
Number of posts	50	20	
Followed by	500	200	
Follows	30	10	
Operation: Please choose 'the same' or 'different'.			

Fig. 3. Example of User Matching Tasks.

(Top) Example of an original task. (Bottom) Example of a task with axis consistency. Crowd workers had to compare two social media accounts with attributes, with some missing data, to determine whether they belonged to the same person. In this example, these two accounts are different.

Task 1		
Instruction: Which product is more cost-effective based on the prices and ratings of the following products?		
Operation: Please choose a more cost-effective product, product A, product B, product C.		
Name	Price(dollar)	Rating(point)
Product A	90	5
Product B	60	5
Product C	60	3
Operation: Please choose a more cost-effective product, product A, product B, product C.		

Task 1			
Instruction: Which product is more cost-effective based on the prices and ratings of the following products?			
Operation: Please choose a more cost-effective product, product A, product B, product C.			
Name	Product A	Product B	Product C
Price(dollar)	90	60	60
Rating(point)	5	5	3
Operation: Please choose a more cost-effective product, product A, product B, product C.			

Fig. 4. Example of Product Comparison Tasks.

(Top) Example of an original task. (Bottom) Example of a task with axis consistency. Crowd workers were asked to choose the most cost-effective product. In this example, product B is most cost-effective because of its lower price and higher rating.

Table 1. Accuracy Rate of Each Audio Task Design (Screen Reader Users).

The accuracy rate for all four audio matrix tasks was high. The subjects understood all four tasks. Of these, axis consistency 2D audio matrix task and naive mapping 1D audio matrix task were performed with the highest accuracy rate. The accuracy rate of these two tasks was 1, whereas that of the naive mapping 2D audio matrix task was 0.94. The axis consistency 1D audio matrix task had the lowest accuracy rate (0.88).

	Audio (2D)	**Audio (1D)**
Naive mapping	0.94	1
Axis consistent	1	0.88

Table 2: Average Time of Each Audio Task Design (Screen Reader Users).

Screen reader users are much more responsive in terms of hearing than normal people. The average time of the axis consistency 1D audio matrix task was the longest (66s), whereas that of the axis consistency 2D audio matrix task was the shortest (28s). The average time for performing the naive mapping 2D audio matrix task was 40 s and that for the naive mapping 1D audio matrix task was 43s.

	Audio (2D)	**Audio (1D)**
Naive mapping	40s	43s
Axis consistent	28s	66s

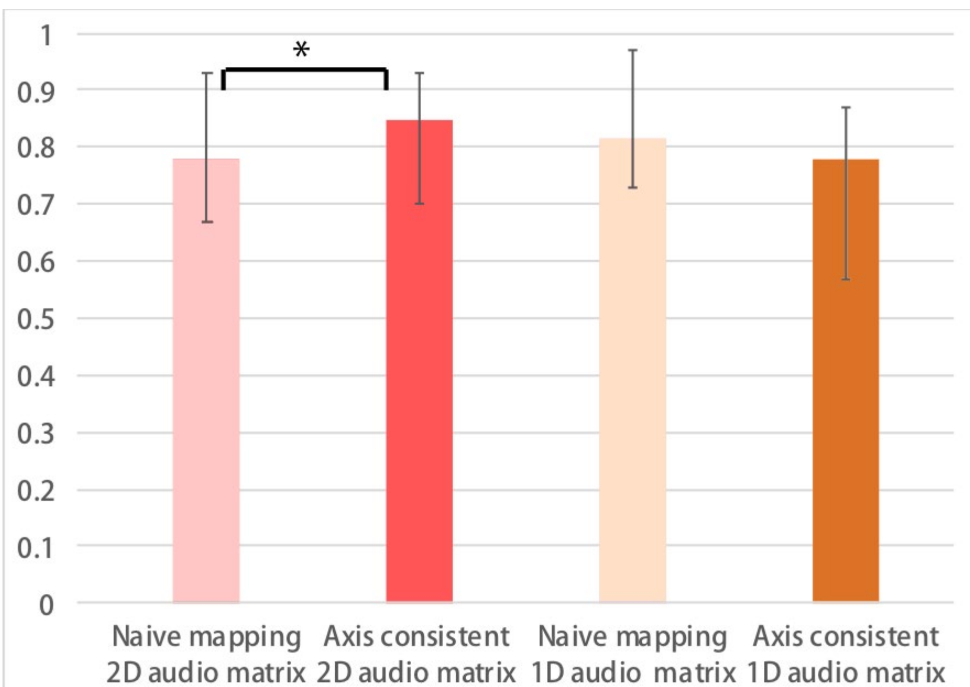

Fig. 5. Accuracy Rate of Four Task Designs.

In audio tasks, the accuracy of the axis consistency 2D audio matrix task is the highest.

Table 3. Average of Total Answer Time of Audio Tasks.

The average time taken in the 1D audio matrix tasks was longer than that in 2D audio matrix tasks. However, the time required for the axis consistency 1D audio matrix task was comparable with other 2D audio matrix tasks, whereas the time required for the naive mapping 1D audio matrix task was much higher.

	Audio (2D)	**Audio(1D)**
Naive mapping	3min 54s	4min 4s
Axis consistent	3min 39s	4min 58s

Works Cited

Abu-hamseh, R., Cordy, J., Andmartin, P. “Schema translation using structural transformation.” CASCON’ 94, vol. 5 (1994): 202–215.

Calvo, R., Kane, S. K., Hurst, A. “Evaluating the accessibility of crowdsourcing tasks on amazon’s mechanical turk.” *Proceedings of ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)* (2014): 257-258.

Christen, P. “The Data Matching Process.” *Part of the Data-Centric Systems and Applications book series (DCSA)* (2012): 23-35.

Jiang, L., Wang Y., Hoffart J., Weikum, G. “Crowdsourced entity markup.” *Proceedings of the 1st international workshop on crowdsourcing the semantic web* (2013): 59-68.

Mudgal, S., Li, H., Rekatsinas, T., Doan, A., Park, Y., Krishnan,G., Deep, R., Arcaute, E., Regavendra, V. “Deep learning for entity matching: A design space exploration.” *Proceedings of the 2018 International Conference on Management of Data* (2018):19-34.

Salisbury, E., Kamar, E., Morris, M. R. “Toward Scalable Social Alt Text: Conversational Crowdsourcing as a Tool for Refining Vision-to-Language Technology for the Blind.” *The 5th AAAI Conference on Human Computation and Crowdsourcing (Hcomp)* (2017).

Sanderson, P. M., Anderson, J., Watson, M. “Extending ecological interface design to auditory displays.” *Proceedings of the 2000 Annual Conference of the Computer-Human Interaction Special Interest Group(CHISIG) of the Ergonomics Society of Australia (OzCHI2000)* (2000): 259-266.

Vashistha, A., Sethi, P., Erson, R. *Bspeak.* "BSpeak: An accessible crowdsourcing marketplace for low-income blind people." *Proceedings of the 2018 CHI Conference on Human Factors in Computing* (2018): 1-13.

Wang, J., Kraska T., Franklin, M. J., Feng, J. "Crowder: Crowdsourcing entity resolution." *Proceedings of the VLDB Endowment (PVLDB)*, Vol. 5, No. 11 (2012): 1483-1494.

Watanabe, T., Yamaguchi, T., Minatani, K. "Advantages and drawbacks of smartphones and tablets for visually impaired people—analysis of ict user survey results." *IEICE TRANSACTIONS on Information and Systems, Vol.E98-D, No.4* (2015): 922-929.

Zhong, Y., Kobayashi, M., Matsubara, M., Morishima, A. "A survey of visually impaired. workers in Japanese and us crowdsourcing platforms." *The 8th AAAI Conference on Human Computation and Crowdsourcing (Hcomp) Work-in-progress session* (2020).

Zhong, Y., Matsubara, M., Kobayashi, M., Morishima A. "Effects of cognitive consistency in microtaskdesign with only auditory information" *Universal Access in Human-Computer Interaction. Applications and Practice. HCII 2020. Lecture Notes in Computer Science*, vol 12189 (2020): 466-476.

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* (2015): 1682–1693.

2021
The Journal on Technology and Persons with Disabilities

Non-Verbal Interaction with Virtual Home Assistants for People with Dysarthria

Aisha Jaddoh, Fernando Loizides, Omer Rana

jaddoha@cardiff.ac.uk , loizidesf@cardiff.ac.uk , ranaof@cardiff.ac.uk

Abstract

This article presents a technique for using non-verbal voice cues to improve interaction with virtual home assistants (VHAs) for people with speech impairments. Non-verbal voice cues are any sounds a person produces other than speech, such as, for example, humming. The architecture of this technique also facilitates the minimization of the privacy and security risks associated with recording and monitoring speech directed at VHAs. The user sends non-verbal voice cues to an intermediary control terminal (Raspberry Pi), which then interprets the sound message and passes on a direct programmatic command to the VHA's cloud service to fulfil the request.

Keywords

accessibility, non-verbal interaction, dysarthria, virtual home assistants, human computer interaction

Introduction

Virtual home assistants (VHAs), such as the Amazon Echo (Tech Advisor), Google Home (Google, "Meet") and Apple HomePod (Apple) are now widely adopted technological devices in the consumer industry. As VHAs are based on automatic speech-recognition systems (ASRs), they are mainly controlled through verbal voice commands. ASRs receive, understand, process, and act upon these commands (Young and Mihailidis 99). Users can ask VHAs to perform certain tasks, such as controlling a home appliance, adding an item to a shopping list, or providing information, such as the daily news. VHAs can provide an important service to people with disabilities. For example, someone with a motor disability could, without physically moving, use a VHA to control their thermostat or to turn off their lights. Similarly, a visually impaired person could listen to the news without the assistance of other technology. However, these devices are not yet fully accessible to people with other, specific types of disability. Communication with VHAs is based on verbal sound commands (i.e., words and sentences), which creates two issues. First, people with speech impairments may have difficulty using VHAs (Takashima et al. 6395)—an effect that increases with the severity of the impairment (De Russis and Corno 163). The reason for this difficulty is found in the nature of their speech, which is different from unimpaired speech. An additional difficulty in collecting data from this group leads to the creation of systems not designed for their speech style (Takashima et al. 6395). Second, VHAs carry privacy and security risks, since they are always listening for specific commands from a user, some of whose data are being transmitted over the internet and stored (Apthorpe et al.).

In our work, we focus on interaction with VHAs for persons with dysarthria, a form of speech impairment. Dysarthria is a complex disorder that causes a disturbed speech process

(Kent et al. 141); it is one of the most severe communication problems, causing some of the highest levels of unintelligible speech (Ansel and Kent 296). It is characterized by poor articulation, difficulty coordinating breaths and speech, and low levels of speech intelligibility. We present a new technique for people affected by dysarthria to interact with VHAs. The proposed technique overcomes dysarthric users' communication difficulties and enables them to use a VHA while simultaneously increasing their privacy and security. In this paper, we present the preliminary work that is required for the improvement in interaction.

This article is organized as follows. As part of the introduction, the related work section demonstrates the research conducted in the field of VHAs and on speech recognition for people with dysarthria. The subsection that follows describes the details of the system architecture. Then, the non-verbal voice subsection addresses the meaning of non-verbal voice cues and the reason behind the selection of the non-verbal voices. The second main section, the discussion, describes the procedure for the system, and then the initial prototype that was implemented. Finally, the concluding section outlines the challenges and future work involved in this approach.

Related Work

HCI research on the field of speech technologies for people with speech impairments is scarce (Ballati et al. 93). A reason for this is the challenges researchers may face when collecting data from people with speech impairments, because of the weakness of their face muscles, which causes fatigue. This results in a limited amount of data (Takashima et al. 6395). Currently, according to the literature, the most common approach to improving interaction with VHAs is to improve their speech-recognition capabilities. Recent studies conducted on VHAs and other voice assistants have reported that the word error rate (WER) is high for dysarthric speakers. In 2019, De Russis and Corno calculated the WER for three different voice assistants used by a group of dysarthric speakers with differing dysarthria severity levels (163). The WER was calculated as an average for all dysarthric users, and as an average for speakers with severe dysarthria only. When all dysarthric speakers were included in the calculation, the average WER for all three voice assistants was 63.4%. For severe speakers, the average for all three voice assistants was 81.9%. Similarly, in 2018, Ballati and others conducted a test on mild-moderate dysarthric speakers and found that the average WER for all users was 44.9 % (93). Mustafa and others reported that speech-to-text recognition accuracy decreases as the severity of dysarthria increases, thus highlighting the strong inverse correlation between speech intelligibility and speech recognition (3924). Researchers have also studied dysarthric people's verbal communication and discovered that it is challenging for a dysarthric person to articulate long sentences. The longer the sentence, the lower the intelligibility and the greater the fatigue experienced by the speaker (Allison et al. 96). This finding supports the idea that using non-verbal voice cues is a beneficial alternative for dysarthric people. Moreover, research reports that some cases with neurological conditions that affect speech will have speech disorders that deteriorate over time (Kent et al. 141). It follows, then, that what works for these people now might be obsolete a while later; by contrast, non-verbal voice cues will be more effective.

Several techniques have been used to improve interaction between people who have speech impairments and voice-controlled devices, such as brain–computer interaction (Luu et al. 1002). Here, brain signals are interpreted as voice commands to control VHAs. Beyond its restrictions of limited technology and capability, this is an intrusive method that is not yet mature enough for commercial use. We aim to provide a non-intrusive approach that replicates the experience people should have when interacting with their VHAs as closely as possible. While scarce work has been undertaken in this area, we can learn from earlier research, which has discussed how people with dysarthria use speech-recognition systems to control the VHS via voice. Two studies (Hawley et al.; Parker et al.) used limited words as a factor for increasing recognition accuracy. However, as some of this restricted collection of words may not be producible by people with sever dysarthria, we take a step further by limiting the vocabulary to a greater extent, to ensure the inclusion of people who cannot pronounce these words, by changing the limited vocabulary to non-verbal speech, which we argue will be more accessible.

Prototype System Architecture

Our prototype employs an architecture and process that enable users to interact with VHAs using non-verbal voice cues. When designing the system, we considered two factors. First, we aimed to facilitate interaction using only non-verbal voice cues, thus allowing users to speak unintelligibly in their interactions with the system. Second, we aimed to ensure users' security and privacy. This work applies edge computing concepts, which involve processing most of the data at the edge of the network rather than in the cloud (Shi et al.). In our prototype's case, the command interpretation is conducted in the Raspberry Pi, which, in addition to reducing the response time, ensures the privacy and security of the data by having a Raspberry Pi serving as an intermediary between the user and the VHA's cloud service. This means that no data are transmitted to the server when words are spoken by the end user. Data that is transmitted and that is likely to be intercepted consists of simple command lines of code, which do not convey any private information, such as the user's voice.

"Non-Verbal Voice"

A non-verbal cue includes any sound a person makes other than speech. Our system has a set of pre-defined non-verbal cues that it can understand, and each cue is mapped to an action. The cues were selected specifically because they are considered to be manageable by broad range of users. The cues were also decided based on dysarthric phonetic features and dysarthric people's articulation capabilities. For example, they account for the fact that a user might inaccurately pronounce consonants or vowels (Kain et al. 743). One study (Lee et al. 34) reported that vowel intelligibility varies according to the severity of a person's dysarthria; the vowel /i/ is less intelligible than the vowel /a/. Dhanalakshmi and others reported that people with dysarthria have difficulty pronouncing mid-vowels, for which the tongue remains in the middle of the mouth while pronouncing the vowel. In Poláček and others, non-verbal voice interaction was divided into three different styles: pitch, where the highness/lowness of the vocal tone was used as a form of interaction (e.g., humming and whistling); vowel sounds; and hissing, which had different applications (862). In our system, the non-verbal cues are as follows:

- a single, non-verbal cue (a different cue length results in a different cue—e.g., "aa" is different from "aaaaa")
- a combination of non-verbal cues from the same sound style (e.g., two different vowels)
- a combination of non-verbal cues from different sound styles (e.g., a vowel and a humming sound).

After deciding the non-verbal list based on the capabilities of people with dysarthria, the commands were chosen from the Amazon Echo and Google websites (Amazon.com; Google, "What"). As there are a variety of commands, we chose the ones most likely to be used every day by people with dysarthria and which could improve their wellbeing. We devised the system language as follows:

- "Ah" translates into "Play the news."
- /æi:/ translates into "What's the weather today?"

- /u:/ translates into "Turn down the volume."
- /u://u:/ translates into "Turn the volume up."
- Two consecutive hums translate into "Play some music."

Discussion

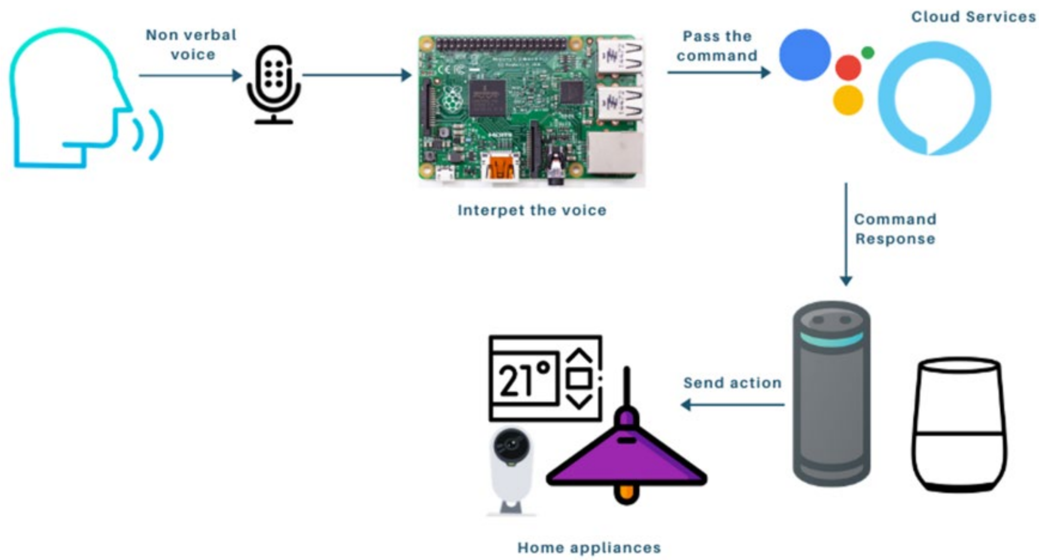

Fig. 1. System Architecture.

The system (see Figure 1) comprises a microphone, Raspberry Pi 3 Model B, the VHA's cloud service (for example, Amazon's web service and Google's assistant service), speakers, and an Internet-of-Things-enabled device that is connected to the VHA via wireless. First, if the user wants, for example, to turn all lights off, they will send a non-verbal voice command to the Raspberry Pi via a microphone. In the Raspberry Pi, the command will be interpreted by signal-processing and machine-learning classification models. The interpreted result will then be passed to the application programming interface (API), that provides access to the VHA cloud service, which processes the request and produces the response. Our prototype uses the Google Assistant Service API, from which the response will be sent back to the appliance (e.g., the light bulbs), and the instruction will be performed. This design has different key features:

- The only physical ability required to use the system is the ability to produce sound.
- The input style (non-verbal) is faster than speech, as, with speech, the system will wait for the utterance to end before performing an action.
- Processing voice only is faster than processing words or sentences (Harada et al.).
- The design is simple and low cost.
- The user's privacy and security are ensured, as the Raspberry Pi serves as an intermediary between the received data and the VHA, and no data aside from the commands made are transmitted to the VHA.

An initial prototype was implemented in python and using conventional neural network (CNN) as a machine-learning technique. CNN is used in sound classifications for different types of sounds.

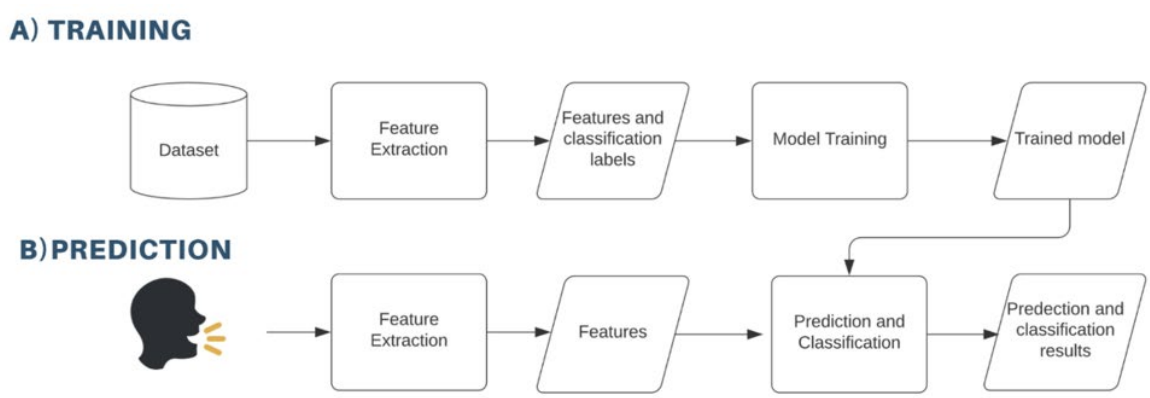

Fig. 2. Voice Classification Flowchart.

The main component of this prototype is the interpretation process, which undergoes two stages (Figure 2): first, the training stage, and second, the prediction stage. In the training stage, the system first extracts the features of the dataset files and transforms the audio files into a dimensional representation using the Librosa library (Librosa).

The extracted features are mel-frequency cepstral coefficients (MFCCs), as they are commonly used in the speech-recognition field and are efficient with different noises (Sahidullah and Saha). In the second stage, after the features of the input command have been extracted, a prediction and classification process will be undertaken to specify for which of the dataset classes this input will be mapped.

The dataset used consists of non-verbal cues from the TORGO database (University of Toronto), a joint effort between the University of Toronto and the Holland Bloorview Kids Rehabilitation Hospital in Toronto. The database contains recordings by eight speakers (three female and five male) with dysarthria and seven speakers (four male and three female) from a non-dysarthric control group. This database was selected because it includes non-verbal recordings. The total number of samples is 36, distributed among three classes.

For the model evaluation, the leave-one-out cross-validation technique (Cawley et al.) was adopted. In this technique, every set in the dataset is used in both training and evaluating the model; one set is taken out as an evaluation set, and the model is trained on the remaining sets. This process is repeated for every set in the training dataset. This technique it is useful for small datasets, such as ours, as it takes advantage of every set.

```
Test accuracy 87.5
Label encoder knows about classes: [0 1 2]
Predicted class 1
0                :  0.00672505935654044151306152343750
1                :  0.66822648048400878906250000000000
2                :  0.32504844665527343750000000000000
```

Fig. 3. Accuracy Results.

Figure 3 shows the results of the prototype. To measure the performance of the model, accuracy metrics were used, and the accuracy rate of this model was 87.5. The performance of this model could be improved with more data.

Conclusion

This work addresses the challenges dysarthric individuals face when using VHAs by introducing a novel interaction technique that allows dysarthric people to interact with these devices through non-verbal voice cues. The design uses a Raspberry Pi as the central part of the system in addition to the VHA. This system will allow dysarthric people to engage with VHA technology independently.

We would like to acknowledge that our study exhibits some limitations at this stage arising from the limited number of datasets available for non-verbal voice cues. Further, there are some challenges for this system. One challenge is scaling limited vocabulary to large number of commands. Moreover, the memorability of the system command is considered a challenge; it could be difficult for the users to remember the list of commands and the actions to which they are mapped. However, this work could be applied to different groups of people who may have difficulty with speech, such as those who suffer from deafness or hearing loss.

Future work will include expanding the system's implementation after collecting dataset recordings from participants (people with dysarthria) and involving them in the design process. For the final stage, a study will be conducted to test the system with persons with dysarthria.

Works Cited

Allison, Kristen M, Yana Yunusova, and Jordan R Green. "Shorter sentence length maximizes intelligibility and speech motor performance in persons with dysarthria due to amyotrophic lateral sclerosis." *American Journal of Speech Language Pathology*, vol. 28, no. 1, 2019, pp. 96–107.

Amazon.com. Accessed 10 October 2020.

Ansel, Beth M, and Raymond D Kent. "Acoustic-phonetic contrasts and intelligibility in the dysarthria associated with mixed cerebral palsy." *Journal of Speech, Language, and Hearing Research*, vol. 35, no. 2, 1992, pp. 296–308.

Apple, "The new sound of home." https://www.apple.com/uk/homepod/. Accessed 21 May 2020.

Apthorpe, Noah, et al. *Spying on the smart home: Privacy attacks and defenses on encrypted IoT traffic*. Cornell University, 2017, https://arxiv.org/abs/1708.05044. Accessed 10 October 2020.

Ballati, Fabio, Fulvio Corno, and Luigi De Russis. "Assessing virtual assistant capabilities with Italian dysarthric speech." *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*. 2018, pp. 93–101.

Cawley, Gavin C., and Nicola LC Talbot. "Efficient leave-one-out cross-validation of kernel fisher discriminant classifiers." *Pattern Recognition* 36.11 (2003): 2585-2592.

De Russis, Luigi, and Fulvio Corno. "On the impact of dysarthric speech on contemporary ASR cloud platforms." *Journal of Reliable Intelligent Environments*, vol. 5 no. 3, 2019, pp. 163–172.

Dhanalakshmi, M, et al. “Speech-input speech-output communication for dysarthric speakers using HMM-based speech recognition and adaptive synthesis system.” *Circuits, Systems, and Signal Processing*, vol. 37, no. 2, 2018, pp. 674–703.

Google, “What you can ask the Google Assistant?” https://support.google.com/googlenest/answer/7172842?hl=en-GB. Accessed 24 June 2020.

Google. “Meet Google Home.” https://store.google.com/gb/product/google home speaker. Accessed 21 May 2020.

Harada, Susumu, Jacob O Wobbrock, and James A Landay. “Voice games: Investigation into the use of non-speech voice input for making computer games more accessible.” *Proceedings of the IFIP Conference on Human-Computer Interaction*. Springer, 2011, pp. 11–29.

Hawley, Mark S, et al. “A speech-controlled environmental control system for people with severe dysarthria.” *Medical Engineering & Physics,* vol. 29, no. 5, 2007, pp. 586–593.

Kain, Alexander B, et al. “Improving the intelligibility of dysarthric speech.” *Speech Communication*, vol. 49, no. 9, 2007, pp. 743–759.

Kent, Ray D, et al. “Acoustic studies of dysarthric speech: Methods, progress, and potential.” *Journal of Communication Disorders,* vol. 32, no. 3, 1999, pp. 141–186.

Lee, Jimin, Emily Dickey, and Zachary Simmons. “Vowel-specific intelligibility and acoustic patterns in individuals with dysarthria secondary to amyotrophic lateral sclerosis.” *Journal of Speech, Language, and Hearing Research*, vol. 62, no. 1, 2019, pp. 34–59.

Librosa. “Librosa Documentation.” https://librosa.org/doc/latest/index.html. Accessed 10 October 2020.

Luu, Brian, et al. "Scalable smart home interface using occipitalis sEMG detection and classification." *Proceedings of the 2018 9th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*. IEEE, 2018, pp. 1002–1008.

Mustafa, Mumtaz Begum, et al. "Exploring the influence of general and specific factors on the recognition accuracy of an ASR system for dysarthric speaker." *Expert Systems with Applications*, vol. 42, no. 8, 2015, pp. 3924–3932.

Parker, Mark, et al. "Automatic speech recognition and training for severely dysarthric users of assistive technology: The STARDUST project." *Clinical Linguistics & Phonetics*, vol. 20, no. 2–3, 2006, pp. 149–156.

Poláček, Ondřej, Zdeněk Mĭkovec, and Pavel Slavĭk. "Predictive scanning keyboard operated by hissing." *Proceedings of the 2nd IASTED International Conference Assistive Technologies*. 2012, pp. 862–869.

Sahidullah, MD and Goutam Saha. "Design, analysis and experimental evaluation of block-based transformation in MFCC computation for speaker recognition." *Speech Communication*, vol. 54, no. 4, 2012, pp. 543–565.

Shi, Weisong, et al. "Edge computing: Vision and challenges." *IEEE Internet of Things Journal*, vol. 3, no. 5, 2016, pp. 637–646.

Takashima, Yuki, Tetsuya Takiguchi, and Yasuo Ariki. "End-to-end dysarthric speech recognition using multiple databases." *Proceedings of the ICASSP 2019–2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* IEEE, 2019, pp. 6395–6399.

Tech Advisor. “What is Amazon Echo? A Complete Guide.” https://www.techadvisor.co.uk/news/audio/amazon-echo-3584881/. Accessed: 21 May 2020.

University of Toronto, Department of Computer Science. “The TORGO Database: Acoustic and articulatory speech from speakers with dysarthria.” http://www.cs.toronto.edu/complingweb/data/TORGO/torgo.html. Accessed 26 May 2020.

Young, Victoria, and Alex Mihailidis. “Difficulties in automatic speech recognition of dysarthric speakers and implications for speech-based applications used by the elderly: A literature review.” *Assistive Technology*, vol. 22, no. 2, 2010, pp. 99–112.

Awareness of Recurrent Education on Visually Impaired Workers: A Questionnaire Survey About Their Working Conditions and Lifelong Learning Courses

Takahiro Miura

National Institute of Advanced Industrial Science and Technology (AIST), Japan.

Yuya Otsuka, Yuri Idesawa, Masatsugu Sakajiri, Junji Onishi

Tsukuba University of Technology, Japan.

Abstract

Although there is a growing need to improve the skills of people with disabilities, it is difficult to establish a scheme through which people with disabilities can learn continuously. In particular, the needs of the visually impaired in recurrent education and lifelong learning are not well-known. Therefore, in this paper, we clarify the needs of recurrent education for visually impaired workers. We conducted a questionnaire survey about the work environment and needs from lifelong learning courses. The results indicated that visually impaired workers generally feel a need to improve their awareness of information technology and communication as well as skills directly related to their work.

Introduction

The lack of human resources has become a problem across the world due to the low birthrates and aging population. Thus, there is a growing need for people with disabilities and older adults to participate in society. Particularly in Japan, with the enactment of the Law for the Elimination of Disability Discrimination, efforts are being made to maintain a certain level of employment for persons with disabilities (i.e., establishing a statutory employment rate). Conversely, although there is a growing need to improve the skills of people with disabilities, it is difficult to establish a scheme through which people with visual and hearing disabilities can learn continuously.

Frameworks for lifelong learning and recurrent education have been established in many countries, introducing recurrent education for people without disabilities (Dennis et al. 1973; Schutze et al. 1987). Recurrent education, as proposed by the Organization for Economic Co-operation and Development (OECD), mainly emphasizes the back-and-forth between educational places such as universities and vocational schools and workplaces (Dennis et al. 1973; Kogan 1979). Alternatively, lifelong education, advocated by the United Nations Educational, Scientific and Cultural Organization (UNESCO), focuses on the continuity of learning throughout the lifetime (vertical articulation to improve coherence) and the linkage of various educational opportunities (horizontal articulation to develop integration between subjects, disciplines or knowledge domains) (Platt 9; Medel-Añonuevo et al. 2001; Elfert 100). In some cases, recurrent education has been introduced for people with disabilities, especially higher education for people with intellectual disabilities (Åkerblom 366). In Japan, Tsukuba University of Technology has begun to introduce a remote recurrent education scheme for the hearing impaired and is developing educational content based on a needs survey (Naito et al. 61). However, the needs of the visually impaired in recurrent education are not well known.

Therefore, this study aims to clarify the needs of visually impaired workers for recurrent education. We prepared and conducted a questionnaire survey about ongoing and continual learning among the visually impaired and obtained answers from 82 working people with visual impairments.

Method

Overview of the questionnaire and procedure

The questionnaire items were determined in discussions among the authors and nonprofit organizations (NPOs) that support visually impaired workers. An overview of the questionnaire items is the following:

- Q1. Basic information (gender, age, degree of disability, educational background)
- Q2. Learning conditions of information subjects when the participant was a student (interest in computers, usage of accessibility hardware/software)
- Q3. Professional career and work conditions (current job, work description, groupware and accessibility hardware/software on the job, what the participant tries to do when working, awareness and use of employment support systems and services)
- Q4. Impressions of higher education and needs for recurrent education (including how well they make use of what they have acquired, what they think they should have been more diligent about, what they needed to learn for their current job, what they want current university students to learn, and what should be enhanced and reduced at universities.)

There were 25 questionnaire categories and 105 items. Questions were written in Japanese. Table 1 shows detailed questionnaire categories.

Table 1. Detailed Questionnaire Categories.

No.	Question
Q1-1	What is your gender?
Q1-2	What is your age?
Q1-3	What is the disability grade of the Physical Disability Certificate issued in Japan?
Q1-4	Please tell us about your visual disability status.
Q1-5	What is the age of onset of visual impairment?
Q1-6	Please tell us the type of junior high school and high school you graduated from, and your final education record.
Q1-7	If you graduated from a university, junior college, or vocational school, please indicate the faculty and department you attended.
Q2-1	When did you first become interested in computers?
Q2-2	Please describe how you became interested in computers and programming, and how you started using them.
Q2-3	On a scale of 1 to 5, please rate your proficiency in computer skills and accessibility features at the time of graduation from your last educational institution.
Q3-1	In what industry do you work now? Choose the one closest to you.
Q3-2-1	Please answer the following questions if you are or were involved in information-related business such as system operation, development, and design. Please select the type of work you have experience with.
Q3-2-2	Please tell us what you work now, if you don’t mind.
Q3-3-1	We would like to ask you if you have any experience in business programming. What languages have you used in the past?
Q3-3-2	Please tell us about your integrated development environment (IDE) for the programming language you selected above.
Q3-4-1	We would like to ask you if you work in an office environment such as clerical work, accounting, or human resources. Do you have any difficulties in using document creation software such as Word or spreadsheet software such as Excel at work?
Q3-4-2	Please indicate what you do to solve the difficulties you have encountered in the above question.
Q3-5	Please select the accessibility features of the support software/hardware you use in your workplace.
Q3-6	What is the purpose of groupware used in the office?
Q3-7	Do you have any difficulties or dissatisfactions in the workplace? Please fill in the following questions if you don’t mind.
Q3-8-1	What do you keep in mind when you work at your workplace? Please choose one that applies to you.
Q3-8-2	Please describe any special consideration you receive from people around you at work.

No.	Question
Q3-9-1	Please choose any system or service that is specific for a person with a disability.
Q3-9-2	Please select the systems and services for employment support specifically for people with disabilities that you have used.
Q4-1	To what extent do you feel that you can make use of what you have learned and acquired at your current workplace? Please rate on a scale of 1 to 5 for each of the following 20 items.
Q4-2	Looking back on your university days, is there anything that you wish you had worked on more diligently? Please rate on a scale of 1 to 5 for each of the following 20 items.
Q4-3	Is there anything new that you needed to learn at your current job? Please select all that apply.
Q4-4	Are there any new things you have learned or are learning at your current job that you should have learned from your university days? Please choose one that applies to you.
Q4-5	Do you have any suggestions for increasing or reducing the educational content for current university students? Please answer with a scale from -5 to -1 if you want to decrease it, or from 1 to 5 if you want to increase it.
Q4-6	Are there any new items that should be reflected in the current information education in universities? If you have any, we would appreciate it if you could describe them.

Participants

Eighty-two (58 males and 24 females) persons with visual impairments participated in this evaluation. The participants' included seven people in their twenties, twenty in their thirties, twenty-seven in their forties, eighteen in their fifties, seven in their sixties, and one above seventy. Five of them were over 65 years old. Twenty-seven participants had total blindness and fifty-five had low vision. In addition, twenty were congenitally visually impaired and fifty-five had acquired visual impairment. The congenital visually impaired participants developed visual impairments when they were equal to or less than six years. Excluding three persons, all participants had a disability certificate that had been issued in Japan.

All participants had graduated from high school, sixty-eight held undergraduate degrees, and seven had attended graduate school. Table 2 shows the participants' occupations. In total, 73 participants answered their current occupations, and most of the participants (N=39, 53%) were clerical workers.

Procedure

The questionnaire was distributed in a computer-based electronic text file format via email. The participants filled out the questionnaire on their own computer and returned it. Approval for the questionnaire survey was obtained from the Tsukuba University of Technology Ethics Board, and all participants had consented to participate in the survey.

Analysis

We analyzed the response trends after a simple tabulation of the responses to each item. Then, to clarify the significant differences among the responses to questionnaire items and to examine the main effects for questionnaire items, analysis of variance (ANOVA) was employed. At that time, the aligned rank transform (ART) (Wobbrock et al. 146) was conducted on the scales prior to ANOVA because the evaluated scales were non-normally distributed. Then, the significance of the main effects was determined using post-hoc multiple comparison methods based on the least-square means and Tukey’s multiplicity adjustment (Lenth 33). We plotted the marked significant differences by boxplots and also displayed the significantly different combinations that indicated by a red rectangle and black asterisk marks. The rectangle is the source, while the asterisks are the comparison, as shown in Fig. 1, 3, 5, and 7.

Next, regarding responses to the questionnaire on some items of impressions of higher education and needs on recurrent education that the participants answered by 5-point Likert scale, multiple ordinal logistic regression (OLR) (McCullagh 127) was employed to examine the significant contribution ratio of each participant attribute, such as gender (Q1-1 in Table 1), age (Q1-2), degrees of disability (Q1-4, and Q1-5), educational background (Q1-6, and Q1-7), and current occupation (Q3-1). Then, each attribute’s log odds ratio was calculated to determine the degree of relevance of the responses. We computed the log odds ratio concerning occupation only regarding whether they were clerical workers, and whether they were professional/technical workers because of the lower number of participants, as shown in Table 2. Subsequently, the effect of the factors was discussed based on the presence or absence of significant differences and the magnitude of the log odds ratio.

Table 2. Participants' Occupations.

Occupation	**Frequency**
Clerical work	39
Professional/Technical	11
Others	9
Sales Professionals	3
Managerial position	3
Health care provider	3
Researcher	2
Studying at vocational school	2
Household/nursing care/barbering	1

Others included as followed:

Occupation	**Frequency**
Sole proprietor Instructor of computer usage Electronic commerce Entertainment (actor/actress) Fortuneteller Planner	5
Unidentified	4

Regarding frequency data, we first performed a simple tabulation. Then, after creating a cross table of frequency data by factors such as disability status, education, and occupation, significant differences were determined by the chi-square test and Cramer's V was calculated as an effect size. After that, we compared Cramer's V with Cohen's criterion (Cohen 159) where there was a significant difference and discussed the effect of each factor.

Results and discussion

Participants' computer skills

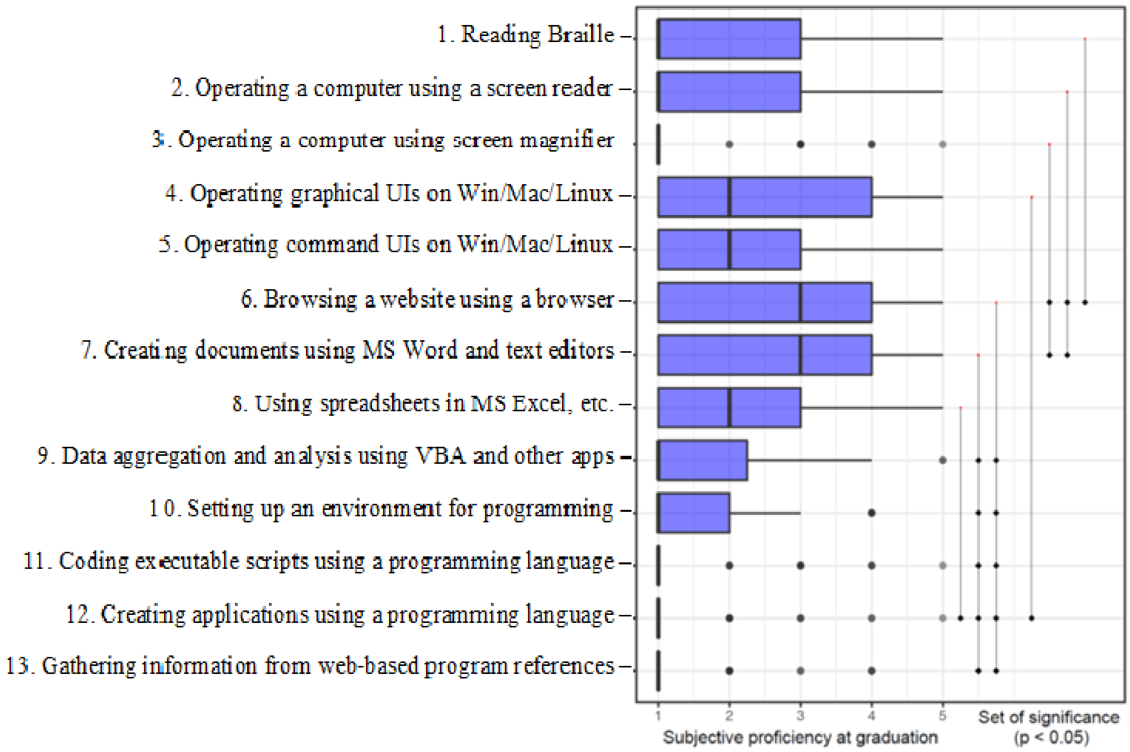

Fig. 1. Subjective Proficiency of the Accessibility Apparatus and Computer Functions when the Participants have Graduated (Right panel of the graph indicates the sets of significant difference).

Figure 1 shows the subjective learning status of the accessibility apparatus and computer functions (Q2-3) in the participants' school days. In particular, they rated their own skills significantly and relatively high for web browsing (No. 6) and text document creation (No. 7), while many gave low ratings for programming skills (No. 9—13). The results of OLR shown in Figure 2 revealed that the blind participants had higher proficiency in braille and screen reading (No. 1—2), but there were significant differences in the responses regarding operating commands on the computer OS, web browsing, and editing text documents and spreadsheets (p<.05).

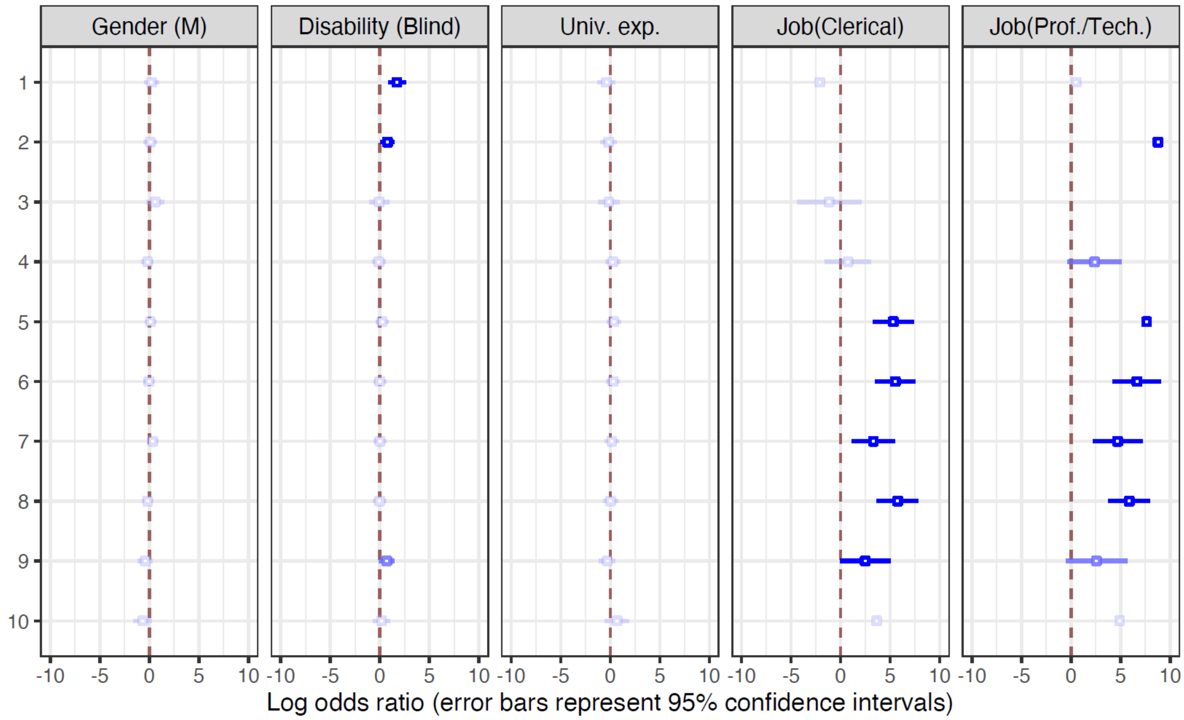

Fig. 2. Log Odds Ratios of Subjective Proficiency Answers Shown in Fig.1 (Q1—Q10) (on the following factors: gender (male or not), disability (blind or not), university experience, job (clerical) or not, and job (professional/technical) or not. The rectangle with error bars in deep blue or middle deep blue represent significant (p<.05) or marginally significant (p<.10) factors, respectively, while those in extremely light blue represent insignificant factors (p>.10))

However, we could not observe significant factors in the programming skills because most of the participants rated too low in the programming-related items. In addition, there were no significant differences among genders (p>.10). In addition, we could not find any difference in computer skills in terms of gender and graduate university experience (p>.10).

Thus, there are likely to be differences in the skills participants attempted to acquire depending on their disability condition and their current occupation. However, most visually impaired workers were not satisfied with the programming skills that they acquired on school days.

Participants' work environments

Table 3. Accessibility Apparatus on the Participants' Workplace and Effect Size.

Accessibility Apparatus	**Frequency**	**%**	**Cramer's V and Significance - Disability (Blind)**	**Cramer's V and Significance - Graduate fr. Univ.**	**Cramer's V and Significance - Job difference**
1. Text-to-Speech function, screen reader	61	74.4	0.35**	0.15	0.21
2. Screen magnification function (zoom-in)	23	28.0	0.32**	0.04	0.29
3. Color and contrast adjustments for the screen	31	37.8	0.39**	0.12	0.35
4. Adjusting the contrast of the shape and color of objects	6	7.3	0.2	0.1	0.3
5. Adjust the size, thickness and color of the text	23	28.0	0.38**	0.1	0.22
6. Adjusting visual effects	10	12.2	0.26*	0	0.48*
7. Gesture support features	3	3.7	0.14	0.1	0.34
8. Devices that display screen information in Braille	13	15.9	0.41***	0.13	0.42
9. Devices that display screen information by dot matrix	2	2.4	0.06	0.12	0.18
10. Braille printers and Braille teplers	4	4.9	0.32*	0.11	0.17
11. Printers for tactile drawings, such as 3D copy machines	2	2.4	0.23	0.12	0.25
12. Others	6	7.3	0	0.1	0.24

***: p<.001, **: p<.01, *: p<.05, +:p<.10

Table 3 shows the accessibility features used by the participants in their work environment (Q3-5 in Table 1). In addition to the various screen readers, they mainly used color adjustment, screen magnification, and screen element size adjustment. There was no significant difference between whether they had a college degree, but there were many significant differences by visual condition (p<.05). Conversely, there was a significant difference in adjusting visual effects between occupations. The reason for this was that two researchers shown in Table 2 chose to adjust the visual effects, while the others did not have much choice. Both researchers had low vision. Thus, this occupational difference was a result of subject bias and not a meaningful difference.

Table 4. Groupware Features Used by Visually Impaired Workers in the Workplace.

Groupware features used by visually impaired workers in the workplace	Frequency	%	Cramer's V and significance - Disability (Blind)	Cramer's V and significance - Graduate fr. Univ.	Cramer's V and significance - Job difference
1. Recording time when arriving and leaving for work	43	52.4	0.15	0.09	0.33
2. Requests on business trip, vacation and travel expense	50	61.0	0.13	0.11	0.4
3. Read the news in the institution	36	43.9	0.11	0	0.34
4. Intra-agency email, chat, and bulletin board	56	68.3	0.02	0	0.4
5. Calendar and schedule management	48	58.5	0.01	0.15	0.31
6. Member search	32	39.0	0.08	0.02	0.39
7. Search and store customer information	11	13.4	0.12	0.17	0.37
8. Contacting the various contacts within the institution	18	22.0	0	0.18	0.34
9. Contact the support office for persons with disabilities	2	2.4	0.11	0.08	0.16
10. Registration for health management and medical checkups	14	17.1	0.03	0.06	0.34

Groupware features used by visually impaired workers in the workplace	Frequency	%	Cramer's V and significance - Disability (Blind)	Cramer's V and significance - Graduate fr. Univ.	Cramer's V and significance - Job difference
11. Booking of rooms such as conference rooms	39	47.6	0.01	0.16	0.33
12. Purchase and procurement of goods	24	29.3	0.17	0.02	0.36
13. Budget management and payment	17	20.7	0.29*	0.02	0.37
14. Creating text, spreadsheet files, etc.	10	12.2	0.1	0.19	0.2
15. Upload and share files	32	39.0	0.03	0.08	0.24
16. Permission requests other than business trips and vacations	35	42.7	0.03	0.05	0.44+
17. Approval of documents	23	28.0	0.09	0.17	0.42
18. Use in other situations	8	9.8	0.03	0.25+	0.44+

***: p<.001, **: p<.01, *: p<.05, +:p<.10

Table 4 shows the functions of the groupware used by visually impaired workers (Q3-6 in Table 1). In addition to communication within the agency, business trip requests and calendar functions were mainly used. Significant differences in visual conditions were found only when they conducted budget management and payments in the system (No. 14). Significant differences between occupations were identified for non-travel permit applications and other situations. It is likely that the applications were unique to their work, as they were professional or research jobs.

Thirty-five users complained of inadequate reading and operation of these groupware products on screen readers, and six users reported that the screen magnification and contrast functions were not working. Three people did not report accessibility problems but did report usability problems, while one person had software limitations due to security issues at their institution. Only six people reported no problems. These results show that about half (53.7%) of the respondents felt that there were accessibility and usability problems in their groupware.

Table 5. What the Participants Try to do While Working at their Workplace.

What the participants try to do when working in the workplace	Frequency	%	Cramer's V and significance - Disability (Blind)	Cramer's V and significance - Graduate fr. Univ.	Cramer's V and significance - Job difference
1. Making efforts to make the situation of visual impairment understandable in the workplace and in the company	61	74.4	0.23+	0.01	0.3
2. Trying to make workers in the industry and others outside of my organization understand the situation of visual impairment	28	34.1	0.21	0.03	0.34
3. Trying to become more proficient in the accessibility apps I use	48	58.5	0.33**	0.04	0.32
4. Preparing several templates for myself to make my work more efficient	36	43.9	0.27*	0.18	0.26
5. Preparing some of my own automation flows (or execution files) to improve business efficiency	16	19.5	0.18	0.09	0.46*
6. The file name and directory structure are designed to retrieve necessary files.	50	61.0	0.29*	0.02	0.31
7. Actively participating in various training and study groups	35	42.7	0.13	0.05	0.3

What the participants try to do when working in the workplace	Frequency	%	Cramer's V and significance - Disability (Blind)	Cramer's V and significance - Graduate fr. Univ.	Cramer's V and significance - Job difference
8. Studying outside of work hours to acquire various certifications	15	18.3	0.06	0.17	0.33
9. Actively communicating on the job to ensure the smooth running of the business	50	61.0	0.19	0.02	0.39
10. Actively engaged in non-work-related communication	35	42.7	0.13	0.07	0.34
11. Actively speaking up to understand the content of the meetings and to make use of my own suggestions.	29	35.4	0.13	0.04	0.3
12. Being proactive in helping surrounding people	37	45.1	0.06	0.11	0.36
13. If there is a difficult task for me, I am proactive in getting others to help me	44	53.7	0.18	0.04	0.4
14. Others	5	6.1	0.07	0.13	0.24

***: p<.001, **: p<.01, *: p<.05, +:p<.10

Table 5 shows what the participants keep in mind when they work at the workplace (Q3-8-1 in Table 1). We found that more than half of the participants tried to make their situation known to the organization they belonged to (No. 1), organized the file/directory structure to improve search efficiency (No. 6), communicated actively with their surroundings (No. 9), learned about the accessibility functions (No. 3), and asked for help when they had problems (No. 13). There was a significant difference in familiarity with accessibility features (No. 3) and the organization of file/directory structures (No. 6) as well as the provision of personal work templates (No. 4) (small to medium effect size). On the other hand, there was a significant difference between occupations in terms of creating an automated flow (No. 5). This item was significantly more common in professional/technical occupations than in other occupations. These results show that visually impaired workers actively communicate with their surroundings to make their own situation known, and they are aware of helping each other when needed. In addition, some participants commented that they asked their colleagues to prepare for accessible files like text formats instead of paper-based documents, presentation files, complex spreadsheets, and image files. On the other hand, they also tend to be devised in the PC environment; in particular, those in technical professions created an automated flow. According to their comments, visually impaired workers should focus on learning a macro language that allows them to control the editor at will.

Impressions of higher education that the participants experienced

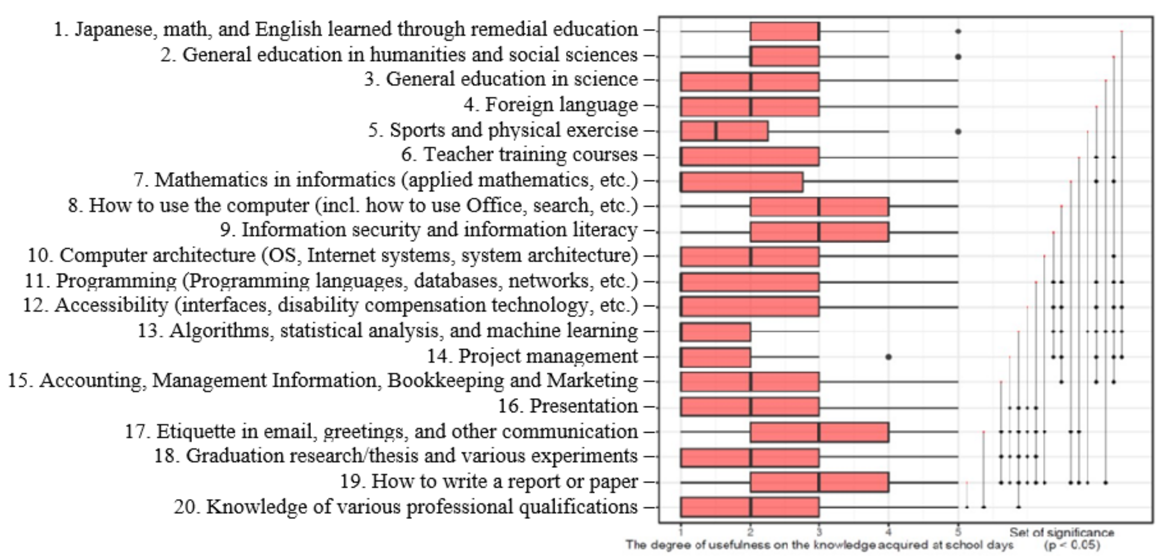

Fig. 3. Degree of Usefulness of the Knowledge Acquired in the Participants' School Days (right panel of the graph indicates the sets of significant difference).

Figure 3 shows the degree of usefulness of the knowledge acquired on school days (Q4-1 in Table 1). In particular, items such as using the computer, information security/literacy, etiquette in email and greetings, and writing a report were rated significantly more useful (p<.05).

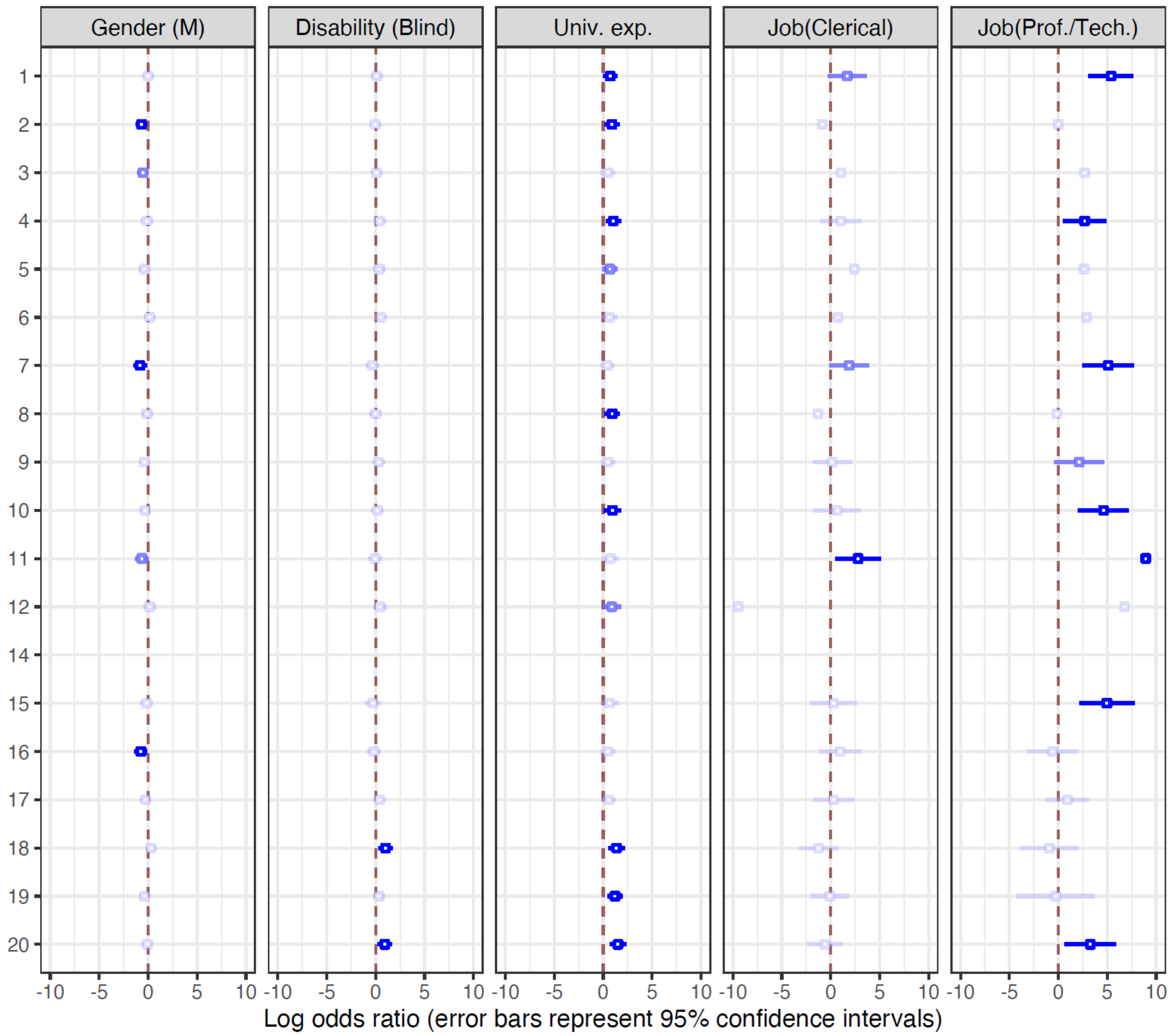

Fig. 4. Log Odds Ratios of Subjective Proficiency Answers Shown in Fig.3 (Q1—Q12, Q14—Q20) (on the following factors: gender (male or not), disability (blind or not), university experience, job (clerical) or not, and job (professional/technical) or not. The log odds ratios to Q13 could not be calculated because of the small number of responses. The rectangle with error

bars in deep blue or middle deep blue represent significant (p<.05) or marginally significant (p<.10) factors, respectively, while those in light blue represent unsignificant factors (p>.10)). The results of OLR shown in Figure 4 revealed that graduates of universities and professional/technical workers significantly emphasized the usefulness of various knowledge acquired at school days. In particular, they scored significantly high in the items of remedial education (No. 1), foreign language (No. 4), computer architecture (No. 10), and knowledge of various professional qualifications (No. 20) (p<.05). University experience was significant for most items. Alternatively, occupational experience tended to increase the log odds ratio for some items. In particular, programming skills (No. 11) scored significantly high in the experience of clerical work and professional/technical work. Therefore, a university or work experience may be a factor in finding what one has learned in the past is significantly useful.

In addition, female participants significantly emphasized the usefulness of humanities and social sciences (No. 2), mathematics in informatics (No. 7), and presentation (No. 16). It is possible that female workers with visual impairments tend to apply what they have learned in the past to the workplace. However, there is still room for reviewing this point by increasing the number of participants.

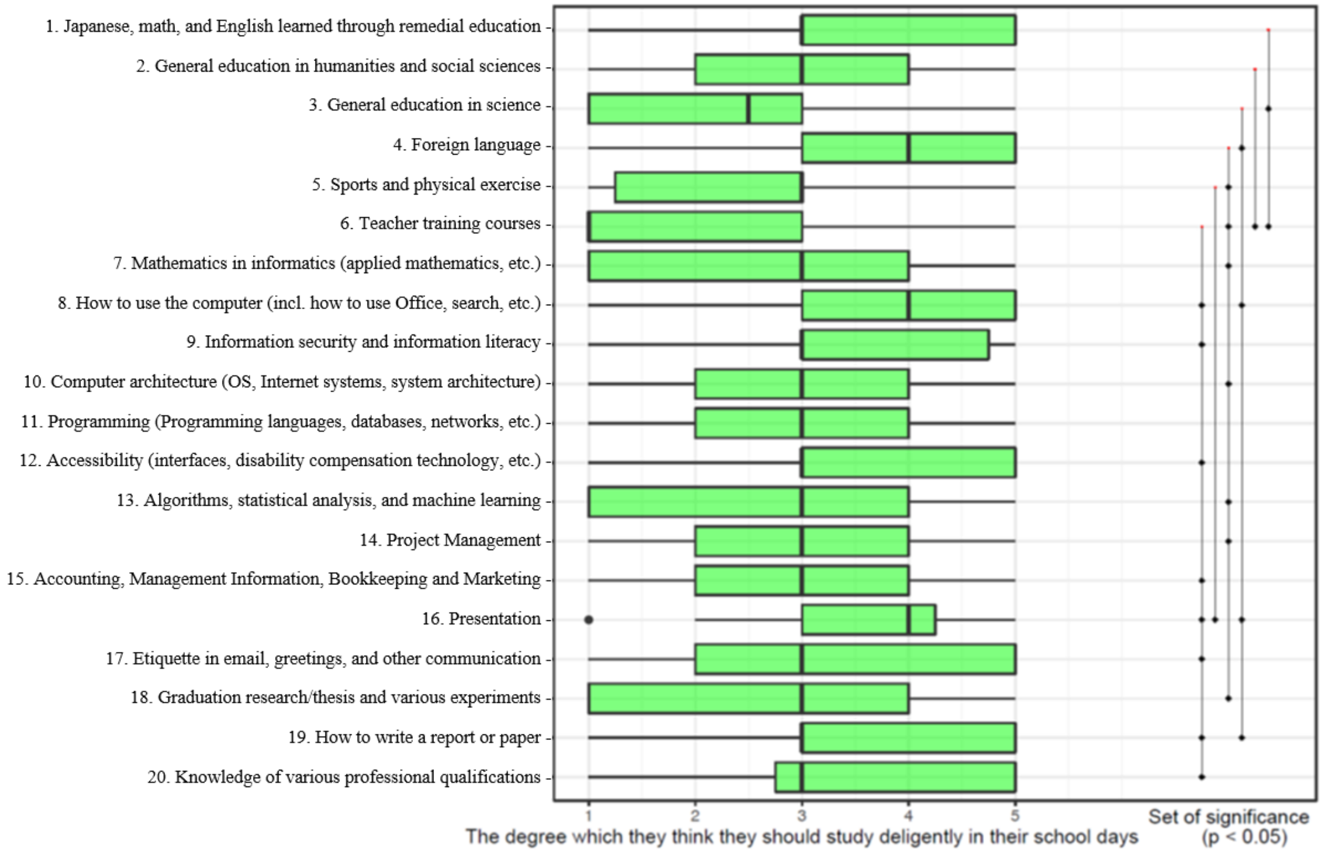

Fig. 5. Degree which the Participants Thought that They Should Study More Diligently in their School Days (right panel of the graph indicates the sets of significant difference).

Figure 5 shows the degree to which they thought that they should study more diligently on their school days (Q4-2 in Table 1). The scores for foreign language, how to use computers, and presentation skills were significantly higher than those of the other items. The results of OLR shown in Figure 6 clarified that those who had a university degree indicated in almost all cases that they should have been more diligent ($p<.05$). This result suggests that many visually impaired people feel the need to relearn these three items, particularly university graduate workers with visual impairments, significantly realize the importance of studying in the past.

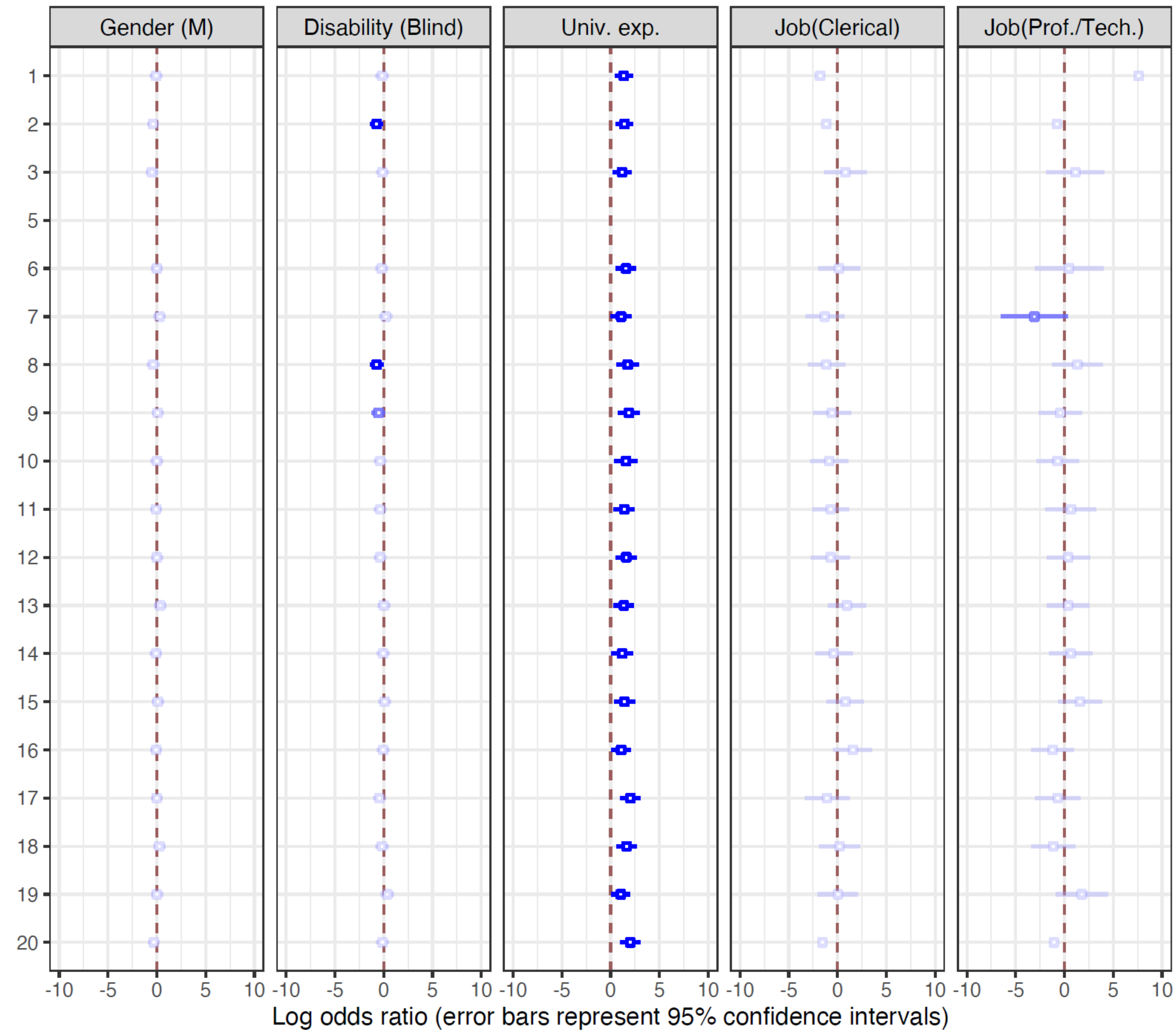

Fig. 6. Log Odds Ratios of the Degree to which the Participants Thought that They Should Study More Diligently in Their School Days Shown in Fig.5 (Q1—Q3, Q5—Q20) (on the following factors: gender (male or not), disability (blind or not), university experience, job (clerical) or not, and job (professional/technical) or not. The log odds ratios to Q4 could not be calculated because of the small number of responses in the participants who did not have university experience. The rectangle with error bars in deep blue or middle deep blue represent significant (p<.05) or marginally significant (p<.10) factors, respectively, while those in light blue represent insignificant factors (p>.10)).

Needs on recurrent education and lifelong learning

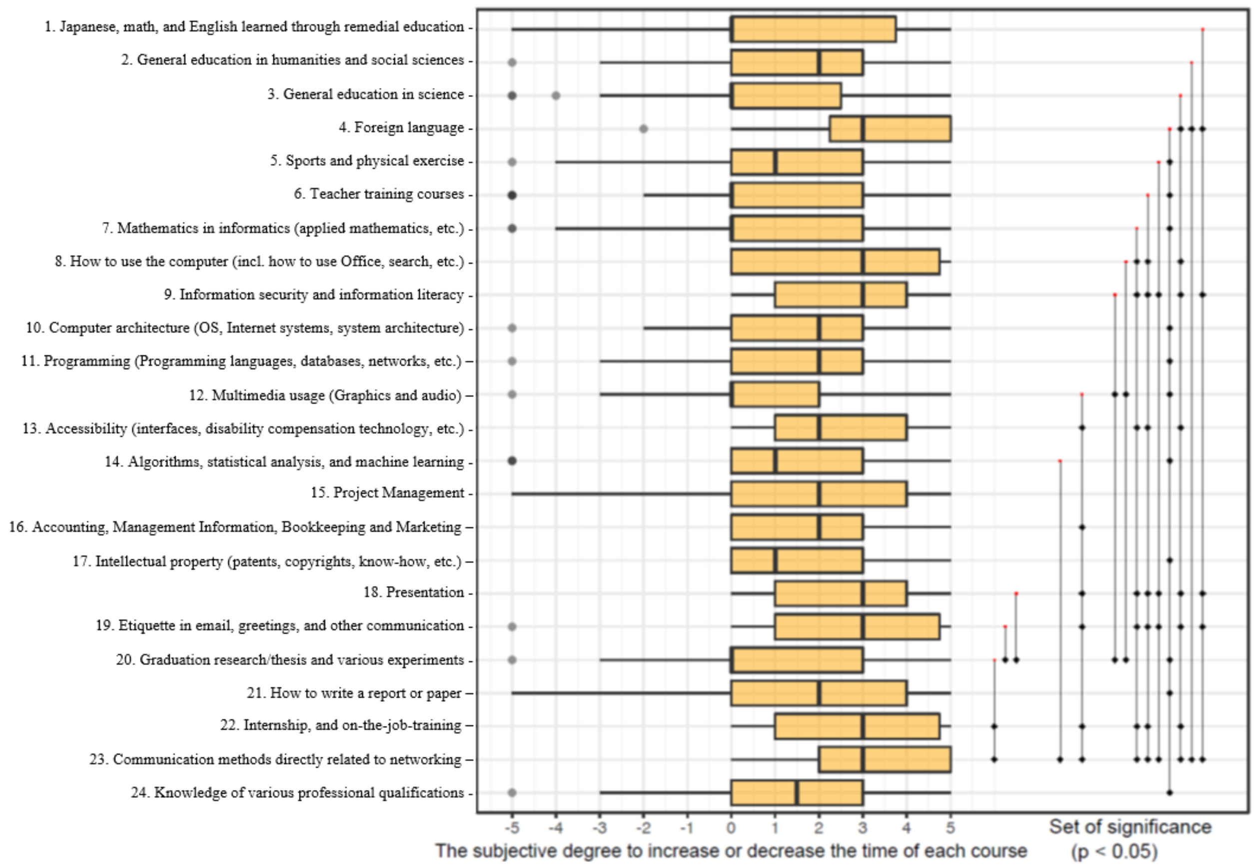

Fig. 7. Subjective Degree to Increase or Decrease the Time of the Academic Courses Held at the Current University Program (-5: decrease, 0: leave unchanged, 5: increase) (right panel of the graph indicates the sets of significant difference).

Figure 7 shows the subjective degree to increase or decrease the time of the academic courses held at the current university program (Q4-5 in Table 1). According to their responses, no items should be reduced, and all courses were rated as either the current state or increased. Notably, their responses indicated that there should be an increasing time of foreign language (No. 4), how to use the computer (No. 8), information security/literacy (No. 9), presentation (No. 18), etiquette in email and greetings (No. 19), and field practices such as internship (No. 22) and job communications (No. 23). In addition, in general, regardless of educational background or current occupation, the participants rated it to be increased in communication-related courses such as No.

18—23. Also, most of the participants who answered Q4-6 in Table 1 mentioned the importance of learning how to use the computer under accessibility software and information security/literacy.

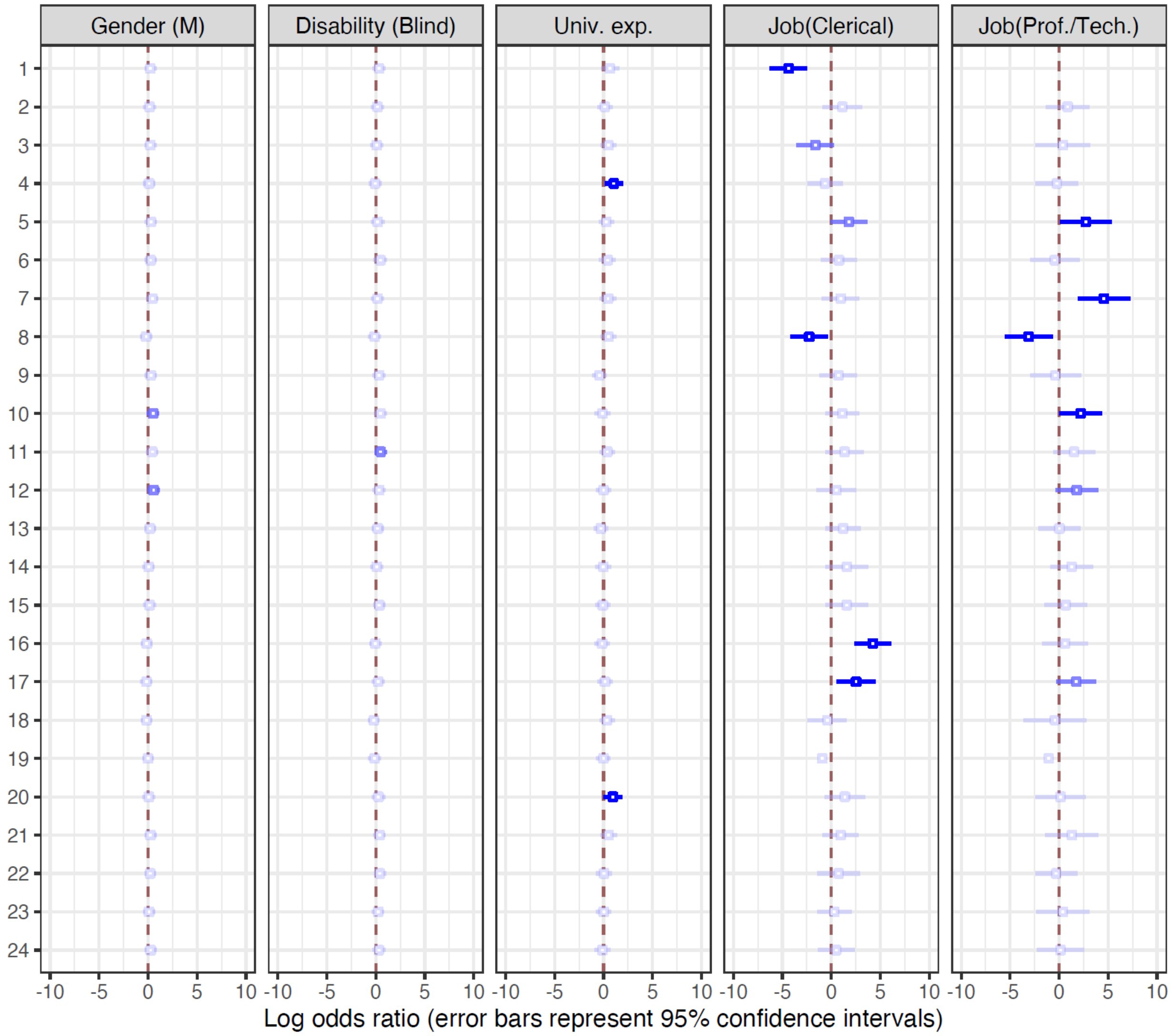

Fig. 8. Log Odds Ratios of the Degree to which Participants Thought They Should Study More Diligently in Their School Days Shown in Fig.7 (Q1—Q3, Q5—Q20) (on the following factors: gender (male or not), disability (blind or not), university experience, job (clerical) or not, and job (professional/technical) or not. The log odds ratios to Q4 could not be calculated because of the small number of responses. The rectangle with error bars in deep blue or middle deep blue

represent significant ($p<.05$) or marginally significant ($p<.10$) factors, respectively, while those in light blue represent unsignificant factors ($p>.10$)).

The results of OLR shown in Figure 8 revealed that gender and disability status did not have a significant effect ($p>.05$) on increasing and decreasing the number of hours in academic courses, but occupation and college education had a significant effect on some items. Notably, visually impaired workers with university degrees significantly answered to increase the time for foreign language (No. 4) and graduation research and various experiments (No. 20). According to one of the participants (answer to Q4-6 in Table 1), visually impaired workers should acquire the skills to develop their ability in their daily life rather than learning at academic lectures. To develop such skills in recurrent education at university, we thought that it would be useful to provide opportunities for the opportunities of project-based learning including graduation research and experiments that make use of what has been learned. However, it is necessary to consider the basic principles of integrated learning in the future.

However, participants with clerical jobs and those engaged in professional/technical work tended to emphasize different courses. Clerical workers mainly increased the time they spent on accounting, management information, bookkeeping, marketing (No. 16), and dealing with intellectual property (No. 17). Alternatively, technical/professional workers significantly increased the time they spent on sports and physical exercise (No. 5), mathematics (No. 7), and computer architectures (No. 10). These results indicate that there is a tendency to increase the number of hours in line with one's current occupation. However, the effect of the current occupation on whether to increase the time spent learning how to use computers (No. 8) was significantly smaller ($p<.05$). This is because visually impaired users have a greater need for computers, regardless of their occupations.

From the above, we can assume that visually impaired workers generally feel a need to improve their awareness of information technology as well as the communication skills directly related to their work. In addition, it became clear that they required knowledge and skills directly related to their jobs in recurrent education.

Conclusion and Future Work

We conducted a questionnaire survey about the work environment and needs of lifelong learning courses for visually impaired workers. Our findings can be summarized as follows:

- To work smoothly, visually impaired workers actively communicate with their surroundings to make their own situation known and help each other when needed.
- Visually impaired workers generally feel a need to improve their awareness of information technology as well as the communication skills directly related to their work.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Numbers 18H01038, 18H01047, and 19K21744, and MEXT Practical Research on Supporting Lifelong Learning for Persons with Disabilities After Graduating from School. We are grateful to members of the authorized nonprofit organization Turtle and the staff of Tsukuba University of Technology for their help.

Works Cited

Åkerblom, Erika. "The will to include–education for becoming a desirable citizen." *Pedagogy, Culture & Society*, 28.3, 2020, pp. 351-366.

Cohen, Jacob. "A power primer." *Psychological bulletin*, 112.1, 1992, pp. 155-159.

Dennis, Kallen, and Jarl Bengtsson. "Recurrent Education: A Strategy for Lifelong Learning." Organisation for Economic Cooperation & Development (OCED), 1973.

Elfert, Maren. "UNESCO, the Faure report, the Delors report, and the political utopia of lifelong learning." *European Journal of Education*, 50.1, 2015, pp. 88-100.

Kogan, Maurice. "Education Policies in Perspective: An Appraisal of OECD Country Educational Policy Reviews." Organisation for Economic Cooperation & Development (OCED), 1979.

Lenth, Russell V. "Least-Squares Means: The R Package lsmeans," *Journal of Statistical Software*, 69.1, 2016, pp. 1-33.

McCullagh, Peter. "Regression models for ordinal data." *Journal of the Royal Statistical Society: Series B (Methodological)*, 42.2, 1980, pp. 109-127.

Medel-Añonuevo, Carolyn, Toshio Ohsako, and Werner Mauch. "Revisiting Lifelong Learning for the 21st Century." UNESCO Institute for Education, 2001.

Naito, Ichiro, Nobuko Kato, Sumihiro Kawano, Hiroshi Murakami, Daisuke Wakatsuki, Tomoyuki Nishioka, Hiroki Minagawa, Hayato Kuroki, Mayumi Shirasawa, Yasushi Ishihara, Shigeki Miyoshi. "Study of Recurrent Education for the Hearing Impaired," *NTUT Techno Report*, 15, 2008, pp. 57-61. (in Japanese)

Platt, William J. "The Faure Report: A Turning Point in Educational Planning." *Educational Planning*, 1, 1, 1973, pp. 1-9.

Schutze, Hans G. and David Istance. "Recurrent Education Revisited: Modes of Participation and Financing." Organisation for Economic Cooperation & Development (OCED), 1987.

Wobbrock, Jacob O., Leah Findlater, Darren Gergle, James J. Higgins. "The Aligned Rank Transform for Nonparametric Factorial Analyses Using Only ANOVA Procedures," *Proc. SIGCHI Conf. Hum. Factors Comput. Syst. (CHI'11)*, 2011, pp. 143-146.

2021
The Journal on Technology and Persons with Disabilities

Haptic Paradigms for Multimodal Interactive Simulations

Jennifer L. Tennison
Saint Louis University
jen.tennison@SLU.edu

Jesse Greenberg
University of Colorado Boulder
jesse.greenberg@Colorado.edu

Emily B. Moore
University of Colorado Boulder
emily.moore@Colorado.edu

Jenna L. Gorlewicz
Saint Louis University
jenna.gorlewicz@SLU.edu

Abstract

Touch is often omitted or viewed as unnecessary in digital learning. Lack of touch feedback limits the accessibility and multimodal capacity of digital educational content. Touchscreens with vibratory, haptic feedback are prevalent, yet this kind of feedback is often under-utilized. This work provides initial investigations into the design, development, and use of vibratory feedback within multimodal, interactive, educational simulations on touchscreen devices by learners with and without visual impairments. The objective of this work is to design and evaluate different haptic paradigms that could support interaction and learning in educational simulations. We investigated the implementation of four haptic paradigms in two physics simulations. Interviews were conducted with eight learners (five sighted learners; three learners with visual impairments) on one simulation and initial results are shared. We discuss the learner outcomes of each paradigm and how they impact design and development moving forward.

Keywords

Blind/Low Vision, K-12 Education, Research & Development

Introduction

Touch is a powerful tool for learning and accessibility (e.g. Klatzky et al.). Touch-based learning experiences are beneficial for all students and are particularly critical for individuals with visual impairments (VI). For digital educational resources, touch-based, haptic technology is essential in advancing inclusive interactive learning. Commercially available touchscreens, such as tablets and smartphones, are uniquely multimodal, capable of displaying visuals, providing audio, and conveying haptic information through vibrations. Effectively leveraging vibrations on mobile devices could benefit learners with diverse needs by adding a haptic modality of interaction with digital content, e.g., graphics.

The role of vibrations on mobile platforms has been used primarily to convey tertiary information, such as alerting users to incoming messages. However, research has shown that vibrations can have broader applications including support learning, navigation, and daily tasks for individuals with VI (Klatzky et al.; Gorlewicz et al.; Giudice et al.). Specifically, vibrations can aid users in understanding static graphics such as basic shapes (Tekli, Issa, and Chbeir; Tennison and Gorlewicz 2016), graphs (Giudice et al.; Palani et al.; Tennison and Gorlewicz 2019), and maps (Poppinga et al.).

Despite its potential benefits for all learners and its availability in commercial hardware, vibratory feedback is not commonly used in educational content. This is in part due to a lack of design guidance on when and how to meaningfully use vibrations in static and dynamic content. In this work, we investigate the use of vibratory haptics in multimodal interactive science simulations. We share outcomes from our initial design and implementation of vibratory haptics for two simulations on mobile devices, and exploratory interviews with sighted and visually impaired learners. Findings indicate challenges and potential next steps for advancing haptics for multimodal interactive learning resources.

Discussion

Simulations

Our learning context included two physics simulations, *John Travoltage* and *Balloons and Static Electricity*, from the collection of PhET Interactive Simulations (PhET) (see *fig. 1*). These simulations were chosen for their open-source code base, and existing auditory display features, including interactive description (auditory description display) (Smith and Moore), and sonifications (the use of non-speech sound to convey information (Tomlinson et al.)). They also represent a comparatively simple (*John Travoltage*) and a more complex interactive experience (*Balloons and Static Electricity*) while addressing the same physics topic (static electricity). Additionally, PhET Interactive Simulations are widely used by teachers and students worldwide. Enhancing these simulations with research-based vibratory haptics has the potential to result in immediate benefits for many learners.

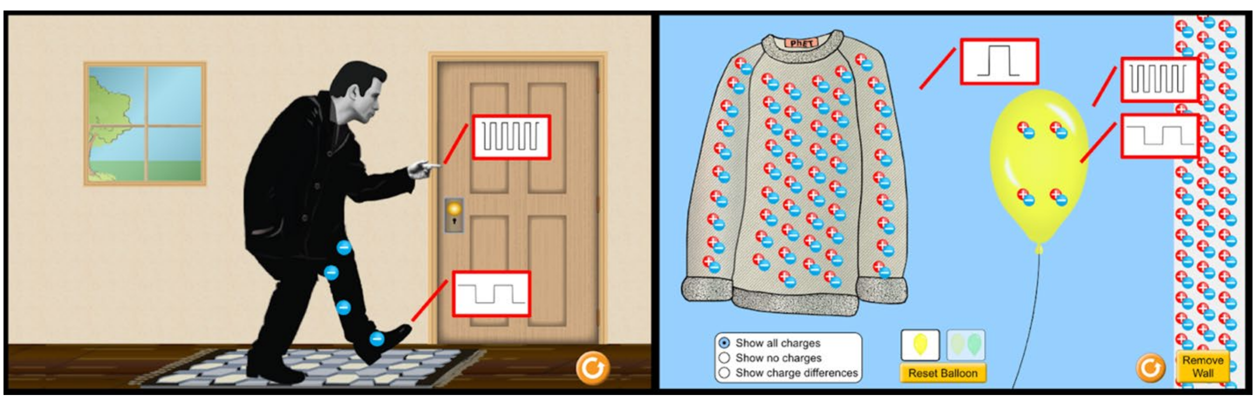

Fig. 1. Examples of Vibratory Implementations for PhET Simulations John Travoltage (left) and Balloons and Static Electricity (right).

Design of Haptic Paradigms

We first designed haptic displays for each simulation. Each physics simulation has a central object (John in *John Travoltage*, and the Balloon in *Balloons and Static Electricity*) akin to a protagonist in a story. There are interactive features of the object the learner can interact with (you can rub John's foot on the rug, and you can move the Balloon onto the sweater or to the wall). Through interaction, the state of the object can change (rubbing John's foot on the rug or the Balloon on the sweater results in a transfer of negative charges onto John or the Balloon, so John or the Balloon can become negatively charged). This change in the object's state can result in changes across the simulation, such as a negatively charged John getting shocked as charges are discharged through a nearby doorknob, and the negatively charged Balloon attracting to - moving across the room to - a positively charged sweater.

Vibration can be used to convey different kinds of feedback, or meaning, to learners, and how learners interpret the vibrations could differ based on the combination of modalities (visual, auditory, and/or haptic) perceivable to the learner simultaneously. To investigate the different kinds of meanings vibrations could convey in the simulations, and how these were perceived and used by learners with and without VI, we designed a set of four haptic paradigms to implement for each simulation. Each paradigm represents one type of meaning that vibrations can convey: 1) Objects; 2) Interactives; 3) Local State changes; and 4) Global State changes (as described in Table 1). The individual paradigms were not intended to represent a complete "final" design; we anticipate that a blend of vibratory feedback from two or more paradigms will ultimately result in the most effective and preferred user experience. The paradigms were to allow for exploration of how learners perceived and made use of each isolated type of vibratory feedback, to provide insights into the potential ways each could be leveraged to achieve an interaction or learning goal of the simulation.

Table 1. Summary of the Four Haptic Paradigms Investigated.

Paradigm	**Description**	***John Travoltage* Implementation with Vibration Parameters**
Objects	Touching the main objects results in vibratory feedback. This paradigm emphasizes the presence of each object, no dynamics are represented by vibrations.	Body ([100, 100], slight pulse) Carpet (Default Vibration) Arm ([25, 25], continuous) Leg ([50, 50], continuous)
Interactives	Moving interactable objects results in vibratory feedback. This paradigm emphasizes interactive objects, no state information is provided when the simulation changes during interaction.	Moving the arm ([25, 25], continuous) Moving the leg ([50, 50], continuous)
Local State Changes	Performing actions which change the simulation results in vibratory feedback. This paradigm emphasizes changes made to the simulation from direct user interaction.	Rubbing the foot against the carpet to generate charge (Default Vibration)
Global State Changes	Highlight changes made to the simulation itself with vibratory feedback. This paradigm emphasizes the resulting state of the simulation after user interaction.	Particles have accumulated on the body ([100, 100], slight pulse) Particles are being discharged to the doorknob ([200, 100], jittery pulse)

Development and Implementation of Haptic Paradigms with Android and iOS

We implemented the haptic paradigms on Android and iOS platforms. We used a Samsung Galaxy Tab S3 tablet (Android) which generates vibration with a coin cell actuator, and an iPhone 11 (iOS) which generates vibration with a linear actuator. Implementing the design paradigms on two platforms allowed for investigation of the differences in affordances related to vibration control, feel, and interaction during use with and without the native screen reader application.

Both platforms allow the developer to control the duration of actuation (which creates the vibration) and the duration of rest (no vibration). Additional options for customizing vibrations on both platforms are available, such as intensity on Android and iOS, and sharpness on iOS (see Weber and Saitis). We used the duration of vibration and rest to generate the vibrations within our four haptic paradigms.

A couple of key differences were discovered between Android and iOS during development. First, Android allows developers web access to vibration triggers inside of browsers, but iOS does not. To trigger vibrations in a custom context on iOS, a dedicated app must be built. Second, the native screen reader for Android, TalkBack, lacks support for the Accessible Rich Internet Applications (ARIA) attribute *valuetext*, which allows for the delivery of text strings (non-numeric descriptions) for the range of sliders, progress bars, and spin buttons.

Challenges were encountered for the combined use of vibratory haptics and screen reader software for both platforms. Potentially most impactful for the design of vibratory haptics supporting access for people with VI is the manner in which screen reader devices handle and intercept touch interactions. For example, freely exploring the onscreen representations through touch, and experiencing vibratory feedback to sense the size and shape of an object is limited, as the touch events are intercepted by the screen reader application and interpreted within its gesture options as taps or swipes, shortening any vibration pattern to the duration of the gesture. This interception occurs when using both TalkBack and VoiceOver, though we found it to be most restrictive using VoiceOver. Additionally, auditory description display implemented in PhET simulations is fully supported for VoiceOver, mobile VoiceOver, NVDA, and JAWS. The lack of support for ARIA *valuetext* decreased the description available to leaners with VI. Because of this, non-visual access to PhET simulations is best using mobile VoiceOver compared to TalkBack, though access to the haptic paradigms was more limited on iOS than Android.

Exploratory Interviews with Learners Using Haptics on Mobile Devices

We conducted semi-structured think-aloud interviews (Lewis, 1982) with visually impaired and sighted learners on an Android tablet with the *John Travoltage* simulation to understand the affordances of each of the paradigms. Five sighted individuals (M = 22.4 years) and three individuals with VI (M = 22.3 years) volunteered to be interviewed (see Table 2 for complete demographics). All learners with VI utilized screen reader software in their daily technology use and did not utilize the visual display during the interviews. All participants were from a midwestern university and compensated with a $25 gift certificate for their time. Interviews took up to one hour to complete. This study was approved by the relevant institutional review board.

Table 2. Participant Demographics.

#	Sex	Age	Visual Impairment
1	Male	22	None
2	Male	23	None
3	Female	21	None
4	Male	25	Retinitis Pigmentosa
5	Female	22	None
6	Male	24	None
7	Male	23	Retinitis Pigmentosa
8	Male	19	Lebers Congenital Amaurosis

The structure of the interviews was similar for all learners, but the interviews with learners with VI were more discussion-based and facilitated than those with sighted learners as all three learners with VI used exclusively iOS, none were familiar with TalkBack. Learners with VI had less access in these exploratory interviews to the vibrations due to standard screen reader operation. We discuss our next steps to mitigate these challenges and increase access to both robust haptic and auditory display in the Conclusions section. For ease of evaluation and comparison of the paradigms, learners with VI could switch between having sonifications and no sonfications as well as switch between paradigms at their own pace. We present findings from both learner groups, with qualitative perceptions being the focus of the learners with VI and performance data being the focus of the sighted learners.

Participants were asked to narrate their experience in exploring each paradigm of each simulation, describing their process in understanding the purpose of the haptics and providing feedback regarding the effectiveness of the haptic renderings. The interviewer observed each learner's interactions and asked clarifying questions, seeking to understand how the learner was exploring, identifying, and interpreting the set of vibrations within each paradigm, as well as their needs and preferences.

During interviews, learners were asked to explore the implementation of each paradigm on *John Travoltage,* both with and without sonifications. For an overview of the auditory description display (used by learners with VI only) and sonifications and their design, see Tomlinson, et al. The interviews with sighted learners were counterbalanced such that half of learners explored all four haptic paradigms on *John Travoltage* without sonifications first, followed by exploring all four haptic paradigms with sonifications.

Results

Learners using visual display and haptic display, with and without sonifications. The five sighted learners were able to complete their first use of the simulation and articulate the main concepts in under a minute (M = 49.48 seconds; *min* = 30 seconds; *max* = 60 seconds). Participants who received the haptic versions of *John Travoltage* with sonifications first (N = 3) were marginally faster (48 seconds, 56 seconds, and 30 seconds) than the individuals who received *John Travoltage* without sonifications first (55 seconds and 60 seconds). All participants confirmed that they were able to feel the vibrations and hear the sonifications well.

Several criteria were used to evaluate the learners' interactions with the corresponding simulation version. These criteria were: 1) identifying the correct number of vibration patterns; 2) articulating the purpose of the vibration; 3) ease of interaction; and 4) finding personal value in the vibrations presented.

When using the Objects paradigm, all learners found value in the vibration patterns and could articulate their purpose, despite not being able to accurately identify all four vibration patterns present. The arm and leg vibrations were the most commonly recognized (with sonifications: 4 of 4 participants; without sonifications: 3 of 4 participants). One learner (in both conditions) discovered the vibration used to represent John's body, but could not articulate it as such. Only one learner (no sonifications) explored the entire on-screen sim area and discovered the rug vibration.

Learners reacted less favorably to the Interactives paradigm. While some learners (with sonifications: 3 of 5 participants; without sonifications: 2 of 5) liked the vibrations associated with moving the arm and leg, all learners across both sonifications and no-sonifications conditions remarked that the vibrations were not meaningful to them. Three of five learners in the sonifications condition indicated that the vibrations were frustrating or redundant with both sonifications and visuals.

The Local State Changes paradigm was the most well-liked haptic paradigm, meeting all learners' expectations for functionality. However, all learners remarked that the simulation felt incomplete with this paradigm. All learners could articulate the purpose of the vibrations (to generate charge with the foot on the rug) and a majority of learners found this to be valuable interaction feedback (with sonifications: 5 of 5 participants; without sonifications: 4 of 5 participants).

The Global State Changes paradigm was also well-liked by learners, but similarly viewed as being an incomplete implementation. Most learners (with sonifications: 4 of 5 participants; without sonifications: 4 of 5 participants) felt this representation was valuable, although the presence of a single continuous vibration pattern to indicate John had accumulated charge was an area of contention. Very few learners liked the implementation of the accumulated charge vibration pattern as implemented (with sonifications: 1 of 5 participants; without sonifications: 2 of 5 participants). All learners initially interpreted this continuous vibration of John's body while having charge as a bug, as all other vibration patterns encountered in the interview were discrete.

Learners using auditory description display and haptic display, with and without sonifications. All three learners with VI could articulate the purpose of the simulation. Enabling sonifications seemed necessary for successful exploration and navigation of the simulation for these learners -- haptics alone was not sufficient. During the first walkthrough of the simulation, two learners requested assistance from the interviewer regarding the operation of TalkBack.

During the exploration of the Objects paradigm, learners did not like the short, choppy vibrations associated with interactive movements. While sighted learners receive continuous vibratory feedback while their finger remains on the model's arm, leg, body and the rug, learners with VI receive abrupt vibrations due to event handling by Talkback. One solution we explored to address this was toggling the screen reader on and off. While turning off the screen reader allowed learners with VI to organically find each component through sound and vibration, learners had difficulty using the interactive components across all haptic paradigms without screen reader assistance. Overall, learners did find value in the proposed solution and desired a way to reconcile the screen reader limitations to receive vibration while interacting with objects.

Similar to sighted learners, learners with VI found the Local State and Global State Changes paradigms to be the most valuable in terms of conveying pedagogical content. However, the learners stressed the importance and the value behind being able to spatially map the simulation itself. All learners with VI desired a version of the simulation that could be "paused" in order to find the various components and spatially explore the content.

Learners with and without sight also shared feedback on specific vibration patterns implemented in the paradigms. Continuous patterns (operating between 100-200 Hz; described as "buzzy") were deemed uninteresting or overstimulating during prolonged exploration. Overwhelmingly, learners enjoyed the pulse vibration implemented to convey the discharge of particles from the finger to the doorknob in the Global State Changes paradigm. The vibration pattern for the foot rubbing against the rug in the Local State Changes paradigm, although "buzzy," was also liked as it met learners' expectations of the interaction. This creates a strong case not just for more varied vibrations, but those that suit the context of the action.

Conclusions

Our goals for this work were to investigate how vibratory haptics can be used to support or enhance interaction with multimodal interactive science simulations. We designed four haptics paradigms for two physics simulations, each paradigm representing one type of feedback (or meaning) the vibrations could convey. We implemented these across two platforms, discovering challenges most notably for the use of auditory description provided through native screen reader software, and access to our custom haptic vibration designs. Overall, all eight participants interviewed indicated that vibratory feedback was an exciting interaction feature and made a meaningful addition to the simulation.

Future Work

From our initial design of haptic paradigms, we will investigate further ways to blend the paradigms to align with the visual and auditory scaffolding of the simulation, supporting initial interaction, and sensemaking with the objects and relationships represented (Podolefsky, Moore, Perkins, 2013). To support learners who could benefit from an exploration of the spatial location of objects with vibratory cues, we will also investigate supporting the addition of a "layer" of vibrations that provide information regarding the objects and interactive elements on-screen. These explorations will include designs with more variations of vibrations, vibration patterns closely aligned with all sonifications, and a systematic exploration of variants of continuous feedback (e.g., fading over time, frequency, etc.). We have also found that the haptics design has interesting overlaps with the sonification design. In the future these two features may pair well together in the design process, and haptic display may benefit from having sound and sonification designers involved in the creation and evaluation of vibration patterns.

The incompatibilities between screen reader applications and vibration capabilities have led us to begin development of self-voicing (with description provided through the browser) with custom gesture control. We plan to investigate this alternative access to haptics vibrations, auditory description display, and sonifications with learners with VI, to inform our haptic designs and potential additional features to complement existing auditory description display provided through screen reader software.

From these investigations, we aim to develop foundational knowledge regarding the perceptual factors that influence effective design of haptic displays, current and emerging possibilities for implementation of haptic displays, and best practice guidelines for the use of haptic displays within interactive learning resources - to create more accessible learning resources for all students.

Works Cited

Giudice, Nicholas A., et al. "Learning non-visual graphical information using a touch-based vibro-audio interface." *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility*. ACM, 2012.

Gorlewicz, Jenna L., et al. "Initial experiences using vibratory touchscreens to display graphical math concepts to students with visual impairments." *Journal of Special Education Technology* 29.2 (2014): 17-25.

Klatzky, Roberta L., et al. "Touch-screen technology for the dynamic display of 2D spatial information without vision: Promise and progress." *Multisensory research* 27.5-6 (2014): 359-378.

Lewis, Clayton. Using the" thinking-aloud" method in cognitive interface design. Yorktown Heights, NY: IBM TJ Watson Research Center, 1982.

Podolefsky, Noah S., Emily B. Moore, and Katherine K. Perkins. "Implicit scaffolding in interactive simulations: Design strategies to support multiple educational goals." arXiv preprint arXiv:1306.6544 (2013).

Palani, Hari Prasath, et al. "Touchscreen-Based Haptic Information Access for Assisting Blind and Visually-Impaired Users: Perceptual Parameters and Design Guidelines." *International Conference on Applied Human Factors and Ergonomics*. Springer, Cham, 2018.

PhET Interactive Simulations, 2020, phet.colorado.edu/.

Poppinga, Benjamin, et al. "TouchOver map: audio-tactile exploration of interactive maps." *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*. ACM, 2011.

Smith, Taliesin L., and Emily B. Moore. "Storytelling to Sensemaking: A Systematic Framework for Designing Auditory Description Display for Interactives." Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020.

Tekli, Joe, Youssef Bou Issa, and Richard Chbeir. "Evaluating touch-screen vibration modality for blind users to access simple shapes and graphics." *International Journal of Human-Computer Studies* 110 (2018): 115-133.

Tennison, Jennifer L., et al. "Establishing Vibration-Based Tactile Line Profiles for Use in Multimodal Graphics." ACM Transactions on Applied Perception (TAP) 17.2 (2020): 1-14.

Tennison, Jennifer L., and Jenna L. Gorlewicz. "Toward non-visual graphics representations on vibratory touchscreens: shape exploration and identification." *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications*. Springer, Cham, 2016.

Tennison, Jennifer L., and Jenna L. Gorlewicz. "Non-visual Perception of Lines on a Multimodal Touchscreen Tablet." *ACM Transactions on Applied Perception (TAP)* 16.1 (2019): 6.

Tomlinson, Brianna J., et al. "A Multimodal Physics Simulation: Design and Evaluation with Diverse Learners." (2019).

Weber, Maximilian and Saitis, Charalampos. "Towards a framework for ubiquitous audio-tactile design." HAID (2020).

2021
The Journal on Technology and Persons with Disabilities

Real-Time Sign Detection for Accessible Indoor Navigation

Seyed Ali Cheraghi, Giovanni Fusco and James M. Coughlan
The Smith-Kettlewell Eye Research Institute
ali.cheraghi@ski.org, giofusco@ski.org, coughlan@ski.org

Abstract

Indoor navigation is a major challenge for people with visual impairments, who often lack access to visual cues such as informational signs, landmarks and structural features that people with normal vision rely on for wayfinding. We describe a new approach to recognizing and analyzing informational signs, such as Exit and restroom signs, in a building. This approach will be incorporated in iNavigate, a smartphone app we are developing, that provides accessible indoor navigation assistance. The app combines a digital map of the environment with computer vision and inertial sensing to estimate the user's location on the map in real time. Our new approach can recognize and analyze any sign from a small number of training images, and multiple types of signs can be processed simultaneously in each video frame. Moreover, in addition to estimating the distance to each detected sign, we can also estimate the approximate sign orientation (indicating if the sign is viewed head-on or obliquely), which improves the localization performance in challenging conditions. We evaluate the performance of our approach on four sign types distributed among multiple floors of an office building.

Keywords

Blindness, Low Vision, Visually Impaired, Accessibility, Navigation, Wayfinding

Introduction

Indoor navigation is a major challenge for people with visual impairments, who often lack access to visual cues such as informational signs, landmarks and structural features that people with normal vision rely on for wayfinding. The most widespread localization approach is GPS, which enables a variety of wayfinding tools such as Google Maps and BlindSquare, but it is only accurate outdoors. Dead reckoning approaches such as step counting using inertial navigation (Flores & Manduchi, 2018) can estimate relative movements indoors or outdoors without any physical infrastructure, but this tracking estimate drifts over time unless it is augmented by absolute location estimates. There are a range of indoor localization approaches, including Bluetooth beacons (Ahmetovic et al., 2016), Wi-Fi triangulation (Heater, 2017) and RFIDs (Ganz et al., 2010). However, all of these approaches incur the cost of installing and maintaining physical infrastructure.

Computer vision is a promising localization approach, but most past work in this area has either required special hardware (Hu et al., 2014) or detailed 3D models of the environment (Gleason et al., 2016) that are time-consuming to generate and make the approach vulnerable to superficial environmental changes (e.g., new carpeting or moved shelves). The iOS app Clew (Yoon et al., 2019) uses visual-inertial odometry (VIO), a function built into modern smartphones combining computer vision and inertial sensing (Kelly & Sukhatme, 2011), to perform dead reckoning. This approach has the advantage of requiring no model of the environment. However, while dead reckoning allows a blind user to retrace their steps from a destination they have already reached back to their starting point, on its own it doesn’t provide guidance to a new destination and does not provide absolute localization.

We developed iNavigate, an accessible localization iPhone app (Fusco & Coughlan, 2020), which combines a 2D map, computer vision and VIO to estimate and track the user's location in an indoor environment. While this approach requires the user to either hold the smartphone or wear it (e.g., on a lanyard) with the camera facing forward while walking, the user doesn't need to aim the camera towards specific signs, which would be challenging for people with low or no vision. We demonstrated the feasibility of our approach with five blind travelers navigating an indoor space, with localization accuracy of roughly 1 meter.

Recently we added verbal turn-by-turn directions to iNavigate (Fusco et al., 2020), thereby creating an accessible wayfinding app that guides the user in real time towards a desired destination. We build on this work by using a more powerful recognition algorithm, YOLOv5, which is able to simultaneously recognize multiple types of signs, with only 8 training images per sign type (instead of the hundreds used before). This will make it easier to create a model for each building, which includes not only the map but also the ability to recognize selected signs inside the building. Moreover, in addition to estimating the distance to each detected sign, we can also estimate the approximate sign orientation (i.e., viewed head-on or obliquely), which can be used to improve the localization performance in challenging conditions.

Discussion

Overview of Sign Detection for Indoor Navigation

The ability to recognize informational signs provides information about the user's location on the map that complements other information sources. For instance, if a specific sign is recognized then the user must be in an area where the sign is visible. If the sign has known physical height and width, we can infer the user's distance from the sign, as well as the approximate orientation of the sign (i.e., viewed head-on or obliquely). This means we can estimate the user's rough location relative to the sign.

Information inferred from sign detections is combined with other information acquired by the iNavigate app, enabling the user's location to be estimated within a meter or better accuracy. This additional information includes visual-inertial odometry (VIO), which fuses computer vision with inertial sensing to estimate relative movements (i.e., dead reckoning), and the locations of walls and other barriers on the map that constrain the estimated trajectory. Note that the fusion of multiple sources of data allows the app to disambiguate which one of multiple identical signs is currently in view.

While iNavigate already uses Exit sign detection as a source of localization information, the new approach we are pursuing provides more localization information. This additional localization information includes the recognition of multiple sign types (see Fig. 1) and estimation of the approximate sign orientation, in addition to the distance to each detected sign. Moreover, whereas the sign recognition algorithms we used in the past required hundreds of sample training images, we demonstrate good recognition results with only eight training images for each type of sign. The need for minimal training data will make it easier to deploy our app in a variety of new buildings, each with its own signs.

Fig. 1. The Four Types of Signs Recognized in this Paper.

Top row: Exit on left and restroom on right. Bottom row: COVID-19 mask on left and fire alarm box on right. The fire alarm box is actually a 3D object but in this paper we treat it as a sign since it is rectangular and nearly flat.

Sign Recognition Algorithm

Previously (Fusco et al., 2020) we used a deep learning model called U-Net (Ronneberger et al., 2015) to detect and segment Exit signs, enabling the distance to be estimated to each sign, with encouraging results. An advantage of U-Net is that it not only recognizes signs but gives precise pixel-by-pixel segmentations. These segmentations allow the contours of the sign to be delineated, which is useful for estimating quantities such as sign distance and orientation. Unfortunately, we have found that U-Net recognition suffers from increased false positive and false negative detections, and inaccurate segmentations, when it is trained on multiple sign types. A separate U-Net could be trained on each sign type, but this approach would be too slow for real-time use for more than just a few sign types.

By contrast, the recently released YOLOv5 object recognizer (Nelson & Solawetz, 2020) is powerful enough to simultaneously recognize many types of signs and runs in real time on a smartphone (several seconds per frame or faster). Our experiments with YOLOv5 show that it is well suited to recognizing a variety of sign types, using only a small number (eight are used in this work) of training images for each type of sign. Note that each training image includes both the target sign of interest, cropped to demarcate a positive example of the sign, and the visual context around the sign (including other objects in the scene), which is used to provide negative examples of imagery to be distinguished from the sign itself. A limitation of YOLOv5 is that it returns an "xy axis-aligned" bounding box (Fig. 2b) around the sign, i.e., a rectangle with sides parallel to the x- and y- axes of the image, instead of a precise pixel-by-pixel segmentation. Fortunately, we have found that the bounding box fits tightly around the rectangular sign, especially since we use the estimated camera roll to undo any camera rotations (Fig. 2a,b) that would make the sign borders appear far from horizontal or vertical.

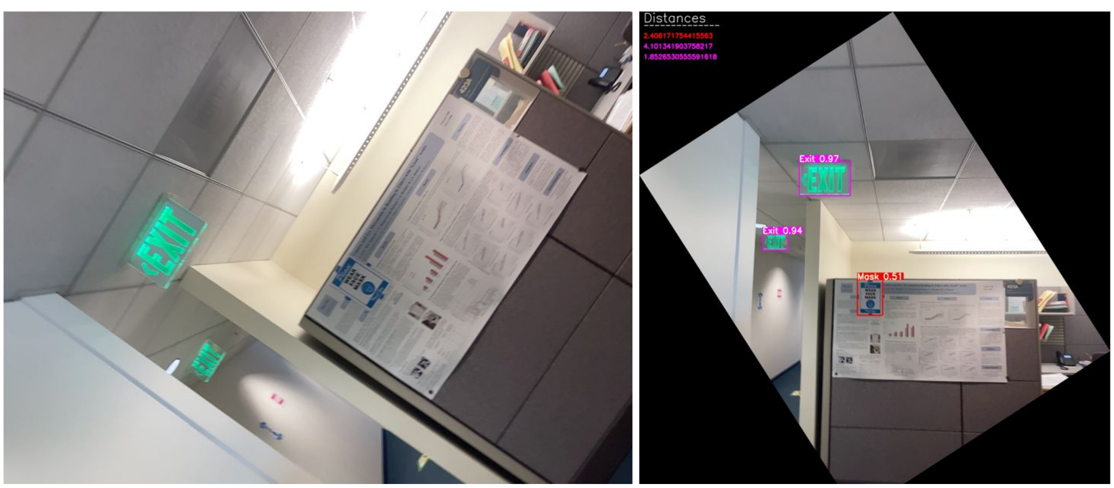

Fig. 2. Sample Image and Sign Detections.

Left to right: (a) Original image taken by iPhone SE shows that the camera is highly "rolled," i.e., horizontal lines in the scene appear far from horizontal in the image. (b) Using the roll angle estimated by the iPhone, the image is unrolled so that horizonal lines appear horizontal. The black triangular regions near the borders correspond to unknown pixels in the unrolled image. YOLOv5 detections are drawn as bounding boxes around one mask sign and two Exit signs. Note that the bounding boxes estimated by YOLOv5 are aligned to the x- and y- axes and fit tightly around the actual signs.

In this work we consider four types of signs to be recognized. The first three are true signs and the fourth is a 3D object that is similar to a sign: the Exit sign, restroom sign, COVID-19 mask sign and a fire alarm box, respectively (Fig. 1). The fire alarm box is a 3D object that is shaped roughly like a small rectangular sign protruding from the wall; we chose this as a “sign” type both because it is an important feature in our building and also to explore how well YOLOv5 works on non-flat objects. We will explore using 3D objects such as water coolers, vending machines, and hand sanitizer stations as visual landmarks in more detail in future work. In the future we will also explore the trade-offs of adding more sign types to YOLOv5, including the trade-off between the rarity of a sign and the amount of localization information it provides.

Distance Estimation

We apply the distance estimation approach we described in (Fusco et al., 2020) to the output of the YOLOv5 algorithm. This approach uses the apparent height of the sign in the image, compared with the sign’s known physical height, to estimate the distance using laws of perspective. It relies on three key assumptions (Fusco et al., 2020):

1. The sign is flat and rectangular, with known physical height (e.g., in cm).
2. It is mounted so that the sign lies in a vertical plane, with the borders of the sign horizontal or vertical with respect to gravity.
3. The camera pitch (angle that the camera line of sight makes with respect to the horizontal plane) and roll (the angle the camera is rotated about its line of sight, with

0° and 90° corresponding to portrait and landscape orientations, respectively) are known.

Fortunately, these assumptions are satisfied for our application. The signs we consider are rectangular, with a standard size, and they are mounted in a way that satisfies assumption 2. (The fire alarm box isn't a true sign, but it is shaped roughly like one.) Moreover, the camera pitch and roll are estimated in real time on modern smartphones using the built-in inertial measurement unit (IMU). Finally, note that the distance estimated by our algorithm is the *straight-line distance along the floor* formed by projecting the 3D camera and sign locations down onto the floor.

Orientation Estimation

A new feature we are exploring is to estimate the sign's orientation to the user by measuring how foreshortened the sign appears in the image. The rough orientation angle is determined by comparing the aspect ratio of the sign in the image with the sign's physical aspect ratio. While approximate, this estimate, when combined with the distance estimate, allows us to estimate the user's rough location relative to the sign. This approach makes the same three key assumptions described in the previous sub-section, Distance Estimation, augmented by knowledge of the physical width of the sign.

More specifically, assuming that the distance to the sign is significantly greater than the height difference between the camera and the sign, the dominant factor that determines the apparent aspect ratio of the sign is the amount of horizontal foreshortening. No foreshortening occurs if the sign is viewed head-on (i.e., the orientation angle θ is 0°). By contrast, significant foreshortening of the apparent width relative to the apparent height (resulting in the bounding box of the sign having a taller, skinnier aspect ratio than the sign's physical aspect ratio) occurs for greater orientation angles (with $\theta \geq 60°$ corresponding to highly oblique viewpoints). We can roughly estimate θ by noting that the amount of foreshortening changes the physical aspect ratio by a factor of $\cos \theta$; this factor can be estimated by comparing the observed aspect ratio of a bounding box detection with the sign's physical aspect ratio. However, since $\cos \theta = \cos (-\theta)$, we can estimate θ only up to an unknown plus or minus sign. This two-fold ambiguity corresponds to the inability to distinguish a sign slanted to the left from a sign slanted an equal angle to the right. (This ambiguity could be resolved for sufficiently close-up signs by observing which of the two vertical sides of the sign appears shorter. However, this approach is unreliable unless the sign is nearby, and we leave this for future work.)

Performance Evaluation on Image Datasets

We acquired three image datasets of the top three floors of the main Smith-Kettlewell building. Because of restrictions due to the COVID-19 pandemic, we were unable to invite visually impaired volunteers to take images. Instead, the experimenter (one of the co-authors) took pictures using an iPhone SE running a data logging app that recorded about 3-5 images per second, along with the pitch, yaw and roll estimated for each image.

The first image dataset was acquired to evaluate the effectiveness of the YOLOv5 object recognition algorithm in terms of *precision* (the proportion of detections that correspond to an actual sign of the corresponding type in the image) and *recall* (the proportion of signs visible in the image that are correctly detected); see (Davis & Goadrich, 2006) for definitions of these measures. For this purpose, the experimenter walked around all three floors of the building while holding the iPhone pointing in the forward direction (in portrait orientation), roughly simulating how a blind person might hold the iPhone while using iNavigate as we observed in our earlier work. The experimental results are shown in Table 1, based on 810 images acquired in total. We note that the recall values are significantly lower than 1.0, but the precision values are close to 1.0. This is appropriate for our planned integration with iNavigate, in which a sign needs only to be correctly recognized in a few frames for it to provide useful localization information; iNavigate can recover from occasional false negative detections.

Table 1. Recall and precision for each class of sign recognition.

Table includes numbers of true positives (TP), false positives (FP) and false negatives (FN), recall and precision. Recall is defined as TP/(TP + FN) and precision as TP/(TP + FP).

Sign type	**TP**	**FP**	**FN**	**Recall**	**Precision**
Exit	458	14	71	0.87	0.97
Restroom	72	0	79	0.48	1.0
Mask	197	5	184	0.52	0.98
Fire alarm	58	2	29	0.67	0.97

The second image dataset, totaling over 8000 images, was acquired to evaluate the distance estimation algorithm. The need for ground truth (actual) distances meant that the experimenter acquired images while standing at rest in multiple locations throughout the top three floors of the building, for which the distances to nearby signs were measured by tape measure. To challenge the distance estimation algorithm, the iPhone was held at multiple angles (e.g., portrait or landscape orientation, upside down, or any roll angle in between), and the hand was sometimes moved to induce the kind of motion blur that often arises in the use of iNavigate in real-world conditions. Given this hand motion, we estimate that the ground truth distances were known to roughly 20 cm accuracy.

To evaluate the performance of our distance estimation algorithm, we estimated the percent distance estimation error, defined as $E=|e-a|/a$ (expressed as a percentage), where e = estimated distance and a = actual distance. The median value of E is reported in Table 2, where it is broken down by sign type and by distance bins (i.e., signs whose actual distance is under a certain threshold are included in the first bin, etc.). These statistics only include distance estimates for signs that are detected, and we note that distance estimates can be distorted if the sign is cut off in the image (which makes it look smaller than it should), or if the bounding box is inaccurately estimated. Because of a bug in the camera logging app, a small number of images had to be discarded because the corresponding roll and pitch values were incorrectly logged. Overall, we find that the median value of *E* is typically under 10% for most signs, with higher values for Exit signs and nearby signs (we are exploring possible explanations for these trends).

Table 2. Median distance estimate error broken down by sign type and actual distance to sign.

Sign type	**Dist < 3 m**	**3 m ≤ Dist < 5 m**	**5 m ≤ Dist < 10 m**	**Dist ≥ 10 m**
Exit	14.3%, 887	12.7%, 586	10.8%, 1129	13.5%, 161
Restroom	5.8%, 340	4.1%, 363	3.9%, 346	N/A
Mask	11.3%, 996	3.7%, 437	3.9%, 453	2.4%, 13
Fire alarm	7.8%, 1303	9.1%, 443	N/A	N/A

Each cell of the table indicates the median percent distance estimation error (see text for details) and the total number of sign detections included in the cell. N/A indicates that no sign detections are available for a cell.

The third image dataset was acquired for a preliminary evaluation of our orientation estimate algorithm. The dataset is small, consisting of just 51 images, with a mask sign and fire alarm box visible in each; the two signs shared the same 2D position (when projected to the floor). The experimenter stood in place at five locations, each a few meters away from the two signs, and aimed the iPhone while holding it in portrait orientation to capture both signs in each photo. The ground truth orientation for each image was either head-on ($\theta_g = 0°$), oblique ($\theta_g = \pm 45°$), or very oblique ($\theta_g = \pm 63°$). The sign orientation estimates are unable to determine whether the orientation angle is positive or negative, so we evaluate the estimation error as follows: $| |\theta_g|$ -

θ|, where θ is the (unsigned, assumed non-negative) orientation estimate. Fig. 3 shows histograms of the estimation error, broken down by sign type and ground truth orientation. The error is fairly low for the mask sign but poor for the fire alarm (except in the head-on case when the orientation was correctly estimated). The mask sign is not only larger than the fire alarm, but more important, it is almost perfectly flat, both of which imply a more accurate prediction of the apparent aspect ratio and thus the orientation. We will measure the orientation accuracy for other sign types in the future, at a range of viewing distances, and will explore possible ways to make the orientation estimate usable for non-flat signs.

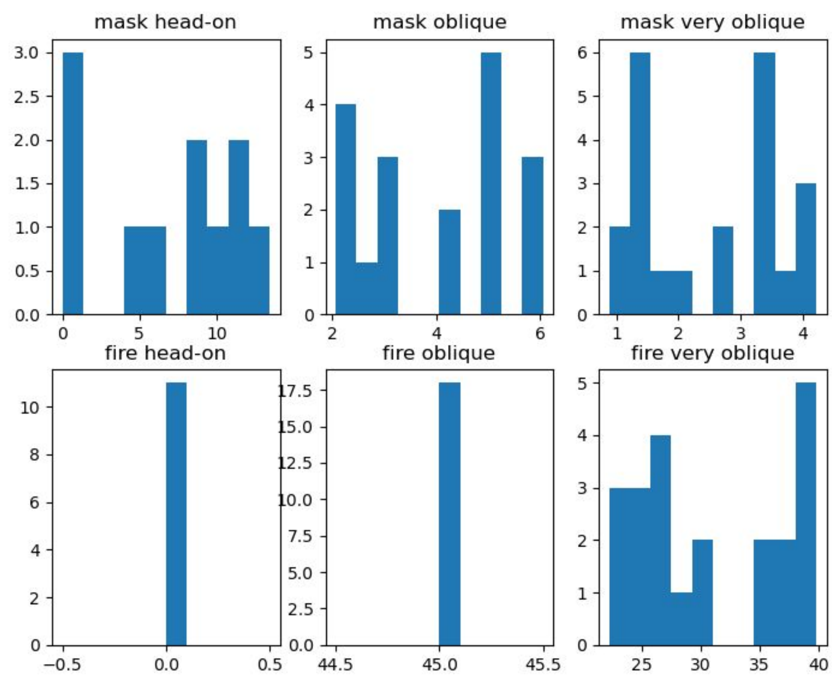

Fig. 3. Orientation Estimation Error Histograms for the Mask Sign and Fire Alarm Box.

Each histogram shows counts on the y-axis as a function of absolute orientation estimation error in degrees on the x-axis. The title of each histogram indicates the sign type (mask on top row and fire alarm on bottom row) and the ground truth orientation of the sign relative to the camera: head-on ($\theta_g = 0°$), oblique ($\theta_g = \pm 45°$) and very oblique ($\theta_g = \pm 63°$). The error is fairly low for the mask sign but poor for the fire alarm (except in the head-on case when the orientation was correctly estimated).

Conclusions

We have demonstrated a new approach to sign detection that is useful for indoor navigation. Our approach allows real-time detection of multiple sign types along with distance and sign orientation estimates that provide useful information about the user's location. Experimental results demonstrate the feasibility of the approach. Our past work with an early version of our wayfinding app, iNavigate, established its usability by blind users. In the future we will integrate our new approach in the app, and we will perform ongoing tests with visually impaired participants as soon as current pandemic restrictions lift.

Acknowledgments

This work was supported by NIH grant 1R01EY029033 and NIDILRR grant 90RE5024-01-00. SAC was supported by Smith-Kettlewell's CV Starr Fellowship.

Works Cited

Ahmetovic, Dragan, et al. "NavCog: turn-by-turn smartphone navigation assistant for people with visual impairments or blindness." *Web for All Conference*. 2016.

Flores, German, and Roberto Manduchi. (2018, April). Easy return: an app for indoor backtracking assistance. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1-12).

Fusco, Giovanni, and James M. Coughlan. "Indoor Localization for Visually Impaired Travelers Using Computer Vision on a Smartphone." *Web for All Conference: Automation for Accessibility*. 2020.

Fusco, Giovanni, et al. "An Indoor Navigation App using Computer Vision and Sign Recognition." *Conference on Computers Helping People with Special Needs*. 2020.

Ganz, Aura, et al. "INSIGHT: RFID and Bluetooth enabled automated space for the blind and visually impaired." *IEEE Engineering in Medicine and Biology*. 2010.

Gleason, Cole, et al. "VizMap: Accessible visual information through crowdsourced map reconstruction." *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*. 2016.

Heater, B. "Apple maps gets indoor mapping for more than 30 airports." TechCrunch.com. 14 December, 2017, https://techcrunch.com/2017/12/14/apple-maps-gets-indoor-mapping-for-more-than-30-airports/ Accessed 17 Nov 2020.

Hu, Feng, Zhigang Zhu, and Jianting Zhang. "Mobile Panoramic Vision for Assisting the Blind via Indexing and Localization." Workshop on Assistive Computer Vision and Robotics. 2014.

Davis, Jesse, and Mark Goadrich. "The relationship between Precision-Recall and ROC curves." *Proceedings of the 23rd International conference on Machine learning*. 2006.

Kelly, Jonathan, and Gaurav S. Sukhatme. "Visual-inertial sensor fusion: Localization, mapping and sensor-to-sensor self-calibration." *The International Journal of Robotics Research* 30.1 (2011): 56-79.

Nelson, J. & Solawetz, J. "YOLOv5 is Here: State-of-the-Art Object Detection at 140 FPS." Roboflow.com. https://blog.roboflow.com/yolov5-is-here Accessed 17 Nov 2020.

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." *International Conference on Medical image computing and computer-assisted intervention*. Springer, Cham, 2015.

Yoon, Chris, et al. "Leveraging Augmented Reality to Create Apps for People with Visual Disabilities: A Case Study in Indoor Navigation." *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*. 2019.

Disability, Technology Based Communication, and Loneliness

Raeda K. Anderson

Shepherd Center, Department of Sociology, Georgia State University

raeda.anderson@shepherd.org

John T. Morris

Shepherd Center

john.morris@shepherd.org

Nicole A. Thompson

Shepherd Center

nicole.thompson@shepherd.org

Ben R. Lippincott

Shepherd Center

ben.lippincott@shepherd.org

Abstract

Examination of the relationship between disability types, technology dependent communication, and loneliness during the COVID19 pandemic using t-tests and regression models on national survey data of adults with disabilities. People with learning disabilities, high levels of worrying, speaking limitations, as well as fatigue and limited stamina report higher levels of loneliness when compared to other people with disabilities. Social interaction with friends as well as living with a spouse are associated with lower levels of loneliness while social interaction via text messages is associated with higher levels of loneliness. Technology dependent communication that mirrors in person interaction, such as video and phone calls, is correlated with lower levels of loneliness while more impersonal interaction, texting, is associated with higher levels of loneliness.

Keywords

Emerging Assistive Technologies, Information & Communications Technology (ICT), Research & Development

Introduction

Having a disability is a chronic stressor and having a disability is associated with poorer mental health across various measures including depression (Travis et al. 265), anxiety (Hendricks et al. 228), stress (Mutkins, Brown, and Thorsteinsson 501), and loneliness (Bauminger, Shulman, and Agam 489). Loneliness is associated with poor physical health factors such as high blood pressure (Hawkley et al. 132), smoking cigarettes (Lauder at al. 233), drinking alcohol (Nieminen et al. 613), and eating disorders (Richardson, Elliot, Roberts 48). Loneliness is also associated with other poor mental health outcomes including depression (Donovan et al. 564), anxiety (White and Roberson-Nay 1010), and deliberate self-harm (Rönkä et al. 4). In addition to poor physical and mental health outcomes, loneliness is associated with a higher rate of mortality (Pantell et al. 2056).

Social interaction serves as a buffer for the effects of a disability on a person's mental health (Tough et al. 294). Most studies on this phenomenon focus on the quality of social interactions and the assistance provided by others, rather than the frequency and communication mode of social interaction (Burholt et al. 1020; Wang et al. 1451). Given that physical distancing requirements and suggestions associated with COVID-19 (Courtemanche et al. 1244), research is needed to examine social interactions for people with disabilities and their mental health in terms of loneliness.

Given the necessity to physically distance ourselves due to COVID19, technology-based communication has become even more integral in facilitating interactions between individuals ranging from phone calls to text messages to video communication. Technology based communication and loneliness has been tested with phone calls (Jin and Park 611), texting (Reidd and Reid 425), online chat (Hu 219), and video calls (Noone et al. 29) with the general population. Notable research has examined technology based communication (Bruner et al. 1028; Caligari et al. 546; Stancliffe et al. 87; Hopps, Pépin, and Boisvert 136), however, the vast majority of research on technology based communication since the beginning of COVID19 focuses on older adults in nursing homes with disabilities (Rorai and Perry 1; Berg-Weger and Morley 1) rather than types of disabilities, social interaction, and technology based communication with loneliness.

Thus, the goal of this study is to examine loneliness across disability types as well as the relationship between disability type, social interaction, and technology dependent communication as a predictor for loneliness for people with disabilities. Using an adapted model from the stress process theory (Pearlin et al. 337) where a stressor causes poor mental health unless buffered by other factors, a partial mediation model outlined in Figure 1 below is used to examine how disability type (stressor) to loneliness (mental health measure) is altered by social interaction and communication modes (buffers).

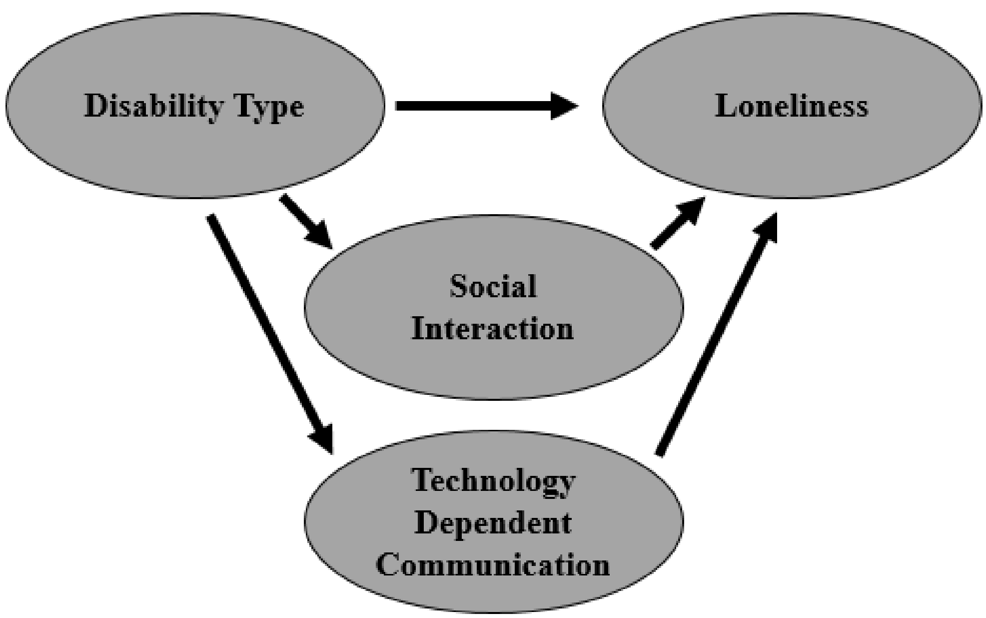

Fig. 1. Conceptual Model. Disability Type, Social Interaction, Technology Dependent Communication, and Loneliness

Data

The data used for this analysis are from the Video Gaming, Gaming, and eSports Survey gathered in September of 2020. The Video Gaming, Gaming, and eSports Survey uses a convenience sample of the Accessibility User Research Collective, AURC, which is a database of 1,020 adults with disabilities across the United States of America. Of these, 455 AURC members completed the survey for a response rate of 44.61%. This online survey is uniquely well suited for the analysis as it contains questions about disability type as well as social interaction patterns, modes of interaction, and a scale measure of loneliness. Listwise deletion was used for analytical purposes, and thus respondents who did not answer all of the study variables outlined below were dropped from the analysis giving an analytical N of 376 respondents.

Loneliness is measured with the UCLA Short Loneliness Scale, a three-question validated measure of loneliness (Vassar and Crosby 602). The UCLA Short Loneliness Scale was developed from a previous iteration that includes twenty indicators (Russel, Peplau, and Ferguson 290). The three indicators that are summated to generate the scale are "*How often do you feel that you lack companionship*?", "*How often do you feel left out*?", and "*How often do you feel isolated from others*?" with response options of (0) hardly ever, (1) some of the time, and (2) often. UCLA Short Loneliness Scale is a summation scale with possible values ranging from 0 to 6 with higher numbers indicating a higher level of loneliness.

Disability is measured with eight dichotomous variables for learning disability, high levels of worrying, difficulty speaking, upper extremity limitations, walking, fatigue and limited stamina, blindness, and deafness. Respondents self-identified for each disability as either having the condition or not having the condition. Most of the disability indicators are measured with a single indicator from the survey including high levels of worrying, difficulty speaking, walking, fatigue and limited stamina, blindness, and deafness. For learning disability, respondents who identified having "*difficulty learning or a learning disability*" and/or "*difficulty concentrating, remembering, or making decisions*" were coded as 1, indicating that they had a learning disability and all respondents were coded as a 0, indicating that they did not have a learning disability. Similarly, for upper extremity limitations, respondents who identified having "*difficulty using your arms*" and/or "*difficulty using your hands and fingers*" were coded as a 1, indicating that they had an upper extremity limitation and all other respondents were coded as a 0, indicating that they did not have an upper extremity limitation.

Social interaction is measured with frequency of initiating conversation, amount of friend interaction, living with a spouse or partner, and gaming as a social activity. Frequency of initiating interaction is measured with a single indicator, "*How frequently do you initiate or start communication in person or via technology with others*?" and response options of (1) never, (2) rarely, (3) sometimes, and (4) often. Amount of friend interaction was measured by the indicator "*How many friends (non-relatives contacted outside business or organizational settings) do you visit, phone, or write to at least once a month?*" with response options of (1) 0 (none), (2) 1-5, (3) 6-10, (4) 11-15, (5) 16-20), and (6) More than 20. Living with a spouse or partner is a dichotomous indicator "*Who do you live with?*" with respondents being coded as a 1 if they indicated that they were living with their spouse or partner, and all other respondents coded as a 0. Respondent playing video games as a social activity was measured with a single indicator, "*Why do you play video games?*" respondents who indicated they play video games as "*a social activity*" were coded as a (1) yes and all other respondents were coded as a (0) no. Social interaction was intentionally measured in a variety of contexts to understand the effects of initiation, quantity, spousal interaction, and social interaction through video games.

Technology dependent communication is measured with dichotomous indicators for video call or webcam, calling on the phone, texting on the phone, and interacting through a video game system. Each of these variables were generated from a check all that apply question of "*How do you usually communicate with others?*" with options for "In-person", "Phone (voice call)", "Phone (text message)", "Video call/webcam", "Though gaming system", and "Other". Given the increased reliance on technology for communication for nearly everyone during COVID19, we opted to analyze both older forms of technology based communication of talking on the phone and texting as well as newer forms of technology communication of video calling and interacting via video game systems.

Control variables for age and sex are used as older adults are more likely to have disabilities (Freedman, Martin, and Schoeni 3138) as well as differences in social interaction patterns (Seeman 243) and females have different social patterns than males (Wong and Heish 703). Age is measured by subtracting year of birth from 2020. Sex is measured from an indicator of "*What is your sex*?" that was recoded into (0) male and other and (1) female.

Analysis

Table 1 includes the descriptive statistics of all study variables. The UCLA short loneliness scale has an average score of 2.56, a standard deviation of 1.96, and ranges from 0. Eight different disability types are measured in this analysis with the largest percent of respondents who have walking limitations (33%), followed by blind (29%), upper extremity limitation (25%), worrying (25%), fatigue and limited stamina (24%), learning disability (20%), speaking (8%), and lastly deaf (5%). Social l Interaction is measured with four indicators;

frequency of initiating communication ranging from 1-4 with 4 being the most interaction (M = 3.31, SD = 0.89), amount of friend interaction ranging from 1-6 with 6 being the most frequent interaction (M = 2.85, SD = 1.44), living with a spouse or partner measured as a dichotomous variable with 47% of respondents as 'yes', and if the respondent plays video games as a social activity measured as a dichotomous variable with 23% of respondents as 'yes'.

Table 1. Descriptive Statistics of Study Variables

Study Variable	**M**	**SD** [a]	**Range**
Loneliness: UCLA Short Loneliness Scale	2.56	1.96	0-6
Disability Type: Learning Disability	0.20	-	0-1
Disability Type: Worrying	0.25	-	0-1
Disability Type: Speaking	0.08	-	0-1
Disability Type: Upper Extremity Limitation	0.25	-	0-1
Disability Type: Walking	0.33	-	0-1
Disability Type: Fatigue & Limited Stamina	0.24	-	0-1
Disability Type: Blind	0.29	-	0-1
Disability Type: Deaf	0.05	-	0-1
Social Interaction: Frequency of Initiating Communication	3.31	0.89	1-4
Social Interaction: Amount of Friend Interaction	2.85	1.44	1-6
Social Interaction: Live with Spouse or Partner	0.47	-	0-1
Social Interaction: Gaming	0.23	-	0-1
Communication Modes: Video Call or Webcam	0.48	-	0-1
Communication Modes: Call (Phone)	0.63	-	0-1
Communication Modes: Text (Phone)	0.86	-	0-1
Communication Modes: Video Game System	0.17	-	0-1
Controls: Age	46.52	14.99	21-86
Controls: Sex (Female)	0.51	-	0-1

[a] standard deviation only reported for continuous variable

N=376

Data: Video Gaming, Gaming, and eSports 2020

Table 2 contains one tail independent sample t-tests examining loneliness levels across specific types of disabilities. These t-tests examine the relationship between each type of disability compared to people who have different disabilities to better understand if different types of disabilities are associated with higher or lower levels of loneliness. Respondents with learning disabilities reported higher levels of loneliness (M = 3.19, SD = 2.26) than people with other types of disabilities (M = 2.42; SD = 1.85); t(374) = -1.904, p = 0.001. Similarly, respondents who worry reported higher levels of loneliness (M = 3.61, SD = 1.96) than people with other types of disabilities (M = 2.23, SD = 1.814); t(374)= -6.200, p < .0001. Respondents with fatigue and limited stamina have higher levels of loneliness (M = 3.09, SD = 2.22) than respondents with other types of disabilities (M = 2.66, SD = 2.00); t(374)= -2.925, p = 0.002. Additionally, respondents who have difficulty speaking have higher levels of loneliness (M = 3.24, SD = 1.92) than respondents who do not worry (M = 2.51, SD = 1.92); t(374) = -1.927, p = 0.028. Thus, respondents with learning disabilities, who worry, have difficulty speaking, as well as those with fatigue and limited stamina report higher levels of loneliness than respondents with different types of disabilities.

Table 2. Disability Type and Loneliness, One Tail Independent Sample T-Test

Disability Type	**Disability?**	**N**	**M**	**SD**	**t**	**sig**
Learning Disability	No Yes	302 74	2.42 3.19	1.85 2.26	-3.006	**
Worrying	No Yes	283 93	2.23 3.61	1.84 1.96	-6.200	***
Speaking	No Yes	347 29	2.51 3.24	1.92 1.94	-1.927	*
Upper Extremity Limitation	No Yes	283 93	3.24 2.56	2.18 1.99	0.057	
Walking	No Yes	251 125	5.60 2.53	1.99 1.94	-0.605	
Fatigue & Limited Stamina	No Yes	285 91	2.66 3.09	2.00 2.22	-2.925	**

Disability Type	Disability?	N	M	SD	t	sig
Blind	No Yes	267 109	2.64 2.40	1.99 1.88	1.045	
Deaf	No Yes	356 20	2.54 3.10	1.94 2.29	-1.244	

*p<.05, **p<.01, *** p<.001

N=376

Data: Video Gaming, Gaming, and eSports 2020

Table 3 presents the findings from the multiple regression analysis in four models. Model 1 tests the relationship between loneliness solely from each of the disability types of learning disability, worrying, speaking, upper extremity limitation, walking, fatigue and limited stamina, blind, and deaf. Respondents who worry report significantly higher levels of loneliness than other types of disabilities ($\beta = 1.11$, $p<.001$).

Model 2 correlates loneliness by facets of social interaction. Respondents with more social interactions with their friends report being less lonely compared to respondents with fewer social interactions with friends ($\beta = -0.16$, $p<.01$). Respondents who are living with their spouse or partner report lower levels of loneliness ($\beta = -0.12$, $p<.05$). Respondents who report playing video games as a social activity have higher levels of loneliness than non-gamers ($\beta = 0.23$, $p < .01$).

Model 3 examines the relationship with technology dependent communication and loneliness. Respondents who do video calls or use web cameras report lower levels of loneliness ($\beta = -0.11$, $p < .05$). Respondents who call other people on the phone regularly report lower levels of loneliness ($\beta = -0.11$, $p < .05$). Respondents who text regularly, however, reported marginally higher levels of loneliness ($\beta = 0.11$, $p < .10$). Thus, when solely testing technology types with loneliness, regular communication that includes voice, with and without video, is associated with being less lonely. While texting, a more indirect form of communication compared to video and voice calls, is associated with higher levels of loneliness.

Model 4 brings together disability type (Model 1), social interaction (Model 2), and technology dependent communication (Model 3) into a large model that allows for examination across these categorizations as they relate to loneliness. Across disability categories, respondents who worry report a higher level of loneliness than people with other types of disabilities (β = 0.21, p<.001). For social interaction, respondents with more friend interaction report lower levels of loneliness (β = -0.11, p < .05). Additionally, respondents who live with a spouse or partner report lower levels of loneliness (β = -0.10, p < .05). Respondents who interact via video games as a social activity report higher levels of loneliness (β = 0.20, p < .01) than non-gamers. Technology dependent communication analysis indicates that respondents who communicate via video call or webcam are less lonely (β = -0.14, p < .05) and respondents who communicate regularly through texting have higher levels of loneliness (β = 0.14, p < .01). The significant relationships from Model 4 mirror those of Model 1, Model 2, and Model 3 with one exception. In Model 3, respondents who call on the phone regularly reported marginally lower levels of loneliness. This slight statistical difference is likely due to statistical power. Given that the N is only 376, there is a limited amount of statistical power for this model given the number of independent and control variables.

Table 3. Disability Type, Social Interaction, & Communication Modes with Loneliness, Regression

Disability Type	**Model 1 β**	**Model 1 sig**	**Model 2 β**	**Model 2 sig**	**Model 3 β**	**Model 3 sig**	**Model 4 β**	**Model 4 sig**
Learning Disability	0.07						0.01	
Worrying	1.11	***					0.21	***
Speaking	0.46						0.05	

Upper Extremity Limitation	0.33						-0.04	
Walking	-0.02						0.01	
Fatigue & Limited Stamina	0.41						0.08	
Blind	-0.14						-0.02	
Deaf	0.48						0.09	

Social Interaction	**Model 1 β**	**Model 1 sig**	**Model 2 β**	**Model 2 sig**	**Model 3 β**	**Model 3 sig**	**Model 4 β**	**Model 4 sig**
Frequency of Initiating Communication			0.12				-0.01	
Amount of Friend Interaction			-0.16	**			-0.11	*
Live with Spouse or Partner			-0.12	*			-0.10	*
Gaming			0.23	**			0.20	**

Technology Dependent Communication	**Model 1 β**	**Model 1 sig**	**Model 2 β**	**Model 2 sig**	**Model 3 β**	**Model 3 sig**	**Model 4 β**	**Model 4 sig**
Video Call or Webcam					-0.11	*	-0.14	*
Call (Phone)					-0.10	ł	-0.03	
Text (Phone)					0.11	*	0.14	**
Video Game System					0.05		-0.07	

Control Variables	**Model 1 β**	**Model 1 sig**	**Model 2 β**	**Model 2 sig**	**Model 3 β**	**Model 3 sig**	**Model 4 β**	**Model 4 sig**
Age	-0.02	*	-0.09	ł	-0.16	**	-0.04	-0.02
Sex (Female)	0.03		0.03		0.04		0.01	0.03
Constant	2.98	***	3.82	***	3.34	***	2.61	***
R^2	0.124	0.098	0.063	0.197				

ł p<.10, *p<.05, **p<.01, *** p<.001

N=376

Data: Video Gaming, Gaming, and eSports 2020

Conclusions

Overall, there is a relationship between types of disabilities, social interaction, technology dependent communication, and loneliness. People who have learning disabilities, worry, difficulty speaking, as well as fatigue and limited stamina are associated with higher levels of loneliness than other types of disabilities. Interestingly, only people who have high levels of worrying are a significant predictor for loneliness when controlling for other types of disabilities. Thus, while having a specific type of disability compared to all types of disabilities are in many cases associated with higher levels of loneliness, when accounting for each individual type of disability the significance of the relationship between a specific type of disability and loneliness is no longer present. For people with disabilities, more interaction with friends as well as living with a significant other is connected with being less lonely. While these relationships are not surprising, the findings support Pearlin's stress process model that social interactions are a buffer between the stressor of disability and the outcome of poor mental health (Pearlin et al. 350).

The most important finding of this analysis is that technology dependent communication that mirrors in person interaction, such as video calls and phone calls, is associated with lower levels of loneliness for people with disabilities. Given the high likelihood that COVID19 and physical distancing measures will continue to force social interaction into a largely virtual format, scholars and developers within the assistive technology and disability arenas should work to address gaps in usage of technology that mirror in person interactions focusing specifically on voice and video options. More specifically, developers should commit to testing and improving technology used for communication for people with cognitive limitations as well as limitations with speech to ensure all people with disabilities are able to engage with others via technology.

Works Cited

Bauminger, Nirit, Cory Shulman, and Galit Agam. "Peer Interaction and Loneliness in High-Functioning Children with Autism." *Journal of Autism and Developmental Disorders* 33.5 (2003): 489-507.

Berg-Weger, M, and J E Morley. "Editorial: Loneliness and Social Isolation in Older Adults during the COVID-19 Pandemic: Implications for Gerontological Social Work." *the Journal of Nutrition, Health & Aging* vol. 24,5 (2020): 456-458

Brunner, Melissa, Bronwyn Hemsley, Leanne Togher, and Stuart Palmer. "Technology and Its Role in Rehabilitation for People with Cognitive-Communication Disability Following a Traumatic Brain Injury (TBI)." *Brain Injury* 31.8 (2017): 1028-1043.

Burholt, Vanessa, Gill Windle, Deborah Morgan, and CFAS Wales Team. "A Social Model of Loneliness: The Roles of Disability, Social Resources, and Cognitive Impairment." *The Gerontologist* 57.6 (2017): 1020-1030.

Courtemanche, Charles, Joseph Garuccio, Anh Le, Joshua Pinkston, and Aaron Yelowitz. "Strong Social Distancing Measures in the United States Reduced the COVID-19 Growth Rate: Study Evaluates the Impact of Social Distancing Measures on the Growth Rate of Confirmed COVID-19 Cases Across the United States." 39.7 *Health Affairs* (2020) 1237-1246.

Freedman, Vicki A., Linda G. Martin, and Robert F. Schoeni. "Recent Trends in Disability and Functioning Among Older Adults in the United States: A Systematic Review." *Jama* 288.24 (2002): 3137-3146.

Hawkley, Louise C., Ronald Thisted, Christopher Masi, and John Cacioppo. “Loneliness Predicts Increased Blood Pressure: 5-Year Cross-Lagged Analyses in Middle-Aged and Older Adults.” *Psychology and Aging* 25.1 (2010): 132.

Hendriks, Sanne M., Jan Spijker, Carmilla Licht, Aartjan Beekman, Florian Hardeveld, Ron de Graaf, Neeltje Batelaan, and Brenda Penninx. “Disability in Anxiety Disorders.” *Journal of Affective Disorders* 166 (2014): 227-233.

Hopps, Sandra Lynn, Michel Pépin, and Jean-Marie Boisvert. “the Effectiveness of Cognitive-Behavioral Group Therapy for Loneliness Via Inter Relaychat Among People with Physical Disabilities.” *Psychotherapy: Theory, Research, Practice, Training* 40.1-2 (2003): 136-147.

Hu, Mu. “Will Online Chat Help Alleviate Mood Loneliness?” *Cyberpsychology & Behavior* 12.2 (2009): 219-223.

Jin, Borae, and Namkee Park. “In-Person Contact Begets Calling and Texting: Interpersonal Motives for Cell Phone Use, Face-To-Face Interaction, and Loneliness.” *Cyberpsychology, Behavior, and Social Networking* 13.6 (2010): 611-618.

Lauder, William, Et Al. “A Comparison of Health Behaviours in Lonely and Non-Lonely Populations.” *Psychology, Health & Medicine* 11.2 (2006): 233-245.

Mutkins, Elizabeth, R. F. Brown, and Einar B. Thorsteinsson. “Stress, Depression, Workplace and Social Supports and Burnout in Intellectual Disability Support Staff.” *Journal of Intellectual Disability Research* 55.5 (2011): 500-510.

Nieminen, Tarja, Et Al. “Social Capital, Health Behaviours and Health: A Population-Based Associational Study.” *BMC Public Health* 13.1 (2013): 613-624.

Noone, Chris, Jenny McSharry, Mike Smalle, Annette Burns, Kerry Dwan, Declan Devane, and Eimear Morrissey. "Video Calls for Reducing Social Isolation and Loneliness in Older People: a Rapid Review." *Cochrane Database of Systematic Reviews* 5 (2020): 1-42.

Pantell, Matthew, David Rehkopf, Douglas Jutte, C. Lenoard Syme, John Balmes, and Nancy Adle. "Social Isolation: A Predictor of Mortality Comparable to Traditional Clinical Risk Factors." *American Journal of Public Health* 103.11 (2013): 2056-2062.

Pearlin, Leonard I., Elizabeth Menaghan, Morton Lieberman, and Joseph Mullan. "the Stress Process." *Journal of Health and Social Behavior* (1981): 337-356.

Richardson, Thomas, Peter Elliott, and Ron Roberts. "Relationship Between Loneliness and Mental Health in Students." *Journal of Public Mental Health* 16.2 (2017): 48-54.

Rönkä, Anna Reetta, Anja Taanila, Markku Korianen, Vappu Sunnari, and Arja Rautio. "Associations of Deliberate Self-Harm with Loneliness, Self-Rated Health and Life Satisfaction in Adolescence: Northern Finland Birth Cohort 1986 Study." *International Journal of Circumpolar Health* 72.1 (2013): 21085.

Rorai, Vanessa, and Tam E. Perry. "An Innovative Telephone Outreach Program to Seniors in Detroit, A City Facing Dire Consequences of COVID-19." *Journal of Gerontological Social Work* (2020): 1-4.

Russell, Dan, Letitia Anne Peplau, and Mary Lund Ferguson. "Developing a Measure of Loneliness." *Journal of Personality Assessment* 42.3 (1978): 290-294.

Seeman, Teresa E., Tina Lusignolo, Marilyn Albert, and Lisa Berkman. "Social Relationships, Social Support, and Patterns of Cognitive Aging in Healthy, High-Functioning Older Adults: Macarthur Studies of Successful Aging." *Health Psychology* 20.4 (2001): 243-255.

Shaw, Lindsay H., and Larry M. Gant. "in Defense of the Internet: The Relationship Between Internet Communication and Depression, Loneliness, Self-Esteem, and Perceived Social Support." *Journal of Obstetrics and Gynecology Canada* 41.10 (2004).

Stancliffe, Roger J., Sheryl Larson, Karen Auerbach, Joshua Engler, Sarah Taub, and K. Charlie Lakin. "Individuals with Intellectual Disabilities and Augmentative and Alternative Communication: Analysis of Survey Data on Uptake of Aided AAC, and Loneliness Experiences." *Augmentative and Alternative Communication* 26.2 (2010): 87-96.

Tough, Hannah, Christine Fekete, Martin Brunkhof, and Johannes Sigrist "Vitality and Mental Health in Disability: Associations with Social Relationships in Persons with Spinal Cord Injury and Their Partners." *Disability and Health Journal* 10.2 (2017): 294-302.

Travis, Linda A., Jeffrey Lyness, Cleveland Shields, Deborah King, and Christopher Cox. "Social Support, Depression, and Functional Disability in Older Adult Primary-Care Patients." *The American Journal of Geriatric Psychiatry* 12.3 (2004): 265-271.

Vassar, Matt, and James W. Crosby. "A Reliability Generalization Study of Coefficient Alpha for the UCLA Loneliness Scale." *Journal of Personality Assessment* 90.6 (2008): 601-607.

Wang, Jingyi, Brynmor Llyod-Evans, Domenico Giacco, Rebecca Forsyth, Cynthia Nebo, Farhana Mann, and Sonia Johnson. "Social Isolation in Mental Health: A Conceptual and Methodological Review." *Social Psychiatry and Psychiatric Epidemiology* 52.12 (2017): 1451-1461.

White, Susan W., and Roxann Roberson-Nay. "Anxiety, Social Deficits, and Loneliness in Youth with Autism Spectrum Disorders." *Journal of Autism and Developmental Disorders* 39.7 (2009): 1006-1013.

Wong, Jaclyn S., and Ning Hsieh. "Functional Status, Cognition, and Social Relationships in Dyadic Perspective." *The Journals of Gerontology: Series B* 74.4 (2019): 703-714.

2021
The Journal on Technology and Persons with Disabilities

Survey of User Needs: eGaming and People with Disabilities

Nicole A. Thompson, Nicholas Ehrhardt, Ben R. Lippincott, Raeda K. Anderson, John T. Morris

Crawford Research Institute, Shepherd Center

nicole.thompson@shepherd.org, nick_ehrhardt@yahoo.com, ben.lippincott@shepherd.org, raeda.anderson@shepherd.org, john.morris@shepherd.org

Abstract

The convergence of several trends in recent years – rapid growth of electronic gaming as both participatory sport and spectator sport in the US and globally, growing sophistication and increasing focus on accessibility by the game-making industry, and rapid social changes caused by the 2020 COVID-19 pandemic – requires better understanding of the landscape of electronic gaming for people with disabilities. This paper analyzes survey research data on participation in electronic gaming by people with disabilities to understand levels of participation and social-psychological effects. Convenience sampling was used to collect data via an online survey from 402 adults with various disabilities. Participants were recruited through the Consumer Advisory Network (CAN) and the Accessibility User Research Collective (AURC), national networks of people with disabilities in the United States developed and maintained by researchers at Shepherd Center, a rehabilitation hospital in Atlanta, Georgia. Data show people with learning disabilities, anxiety and limited stamina play electronic games significantly more than those with other disabilities. Word games are the ones most played by each disability type, followed by role-playing, action and adventure games. Respondents play mostly alone and for fun and relaxation. Small percentages of respondents play e-games for building skills, exercise, learning or rehabilitation.

Keywords

Information & communications technology (ICT), video games, electronic games, e-gaming, disabilities

Introduction

Participation in electronic gaming (or e-gaming, e-sports) has grown considerably in the United States and globally in recent years and is projected grow rapidly in coming years. The Consumer Technology Association's survey titled "2019 Future of Gaming" estimated that 192 million people in the United States age 13-64 (70% of Americans in that age range) played electronic games (2019). Consumers in the United States spent an estimated $43.3 billion on e-gaming in 2018, double the $21.5 billion spent just 5 years earlier in 2013 (Statista, 2020).

Additionally, viewing e-sports competitions online has grown to exceed traditional entertainment such as television, movies, and music. Global e-sports viewership is expected to grow at a 9% compounded annual growth rate (CAGR) between 2019 and 2023, from 454 million viewers in 2019 to 646 million in 2023. The growth of the e-sports viewing audience is on pace to nearly double over the six-year period from the 335 million viewers in 2017. A separate study estimates 2019 global revenue for spectator e-sports at $196 billion (Jones, 2020), a figure that exceeds the combined revenues of video streaming on demand (e.g., Netflix, Hulu) at $50.2 billion (Statista, 2020), global movie box office revenue at $43 billion (Jones, 2020) and global recorded music at $20.2 billion (IFPI, 2020).

Games – electronic or otherwise – can offer substantial social and intellectual development benefits to players, including: forming new friendships for emotionally shy individuals (Kowert, Domahidi and Quandt, 2014); development of moral reasoning (Hodge, Taylor and McAlany, 2020); recovery from lower back pain (France and Thomas, 2018); improvement in motor function for individuals with cerebral palsy (improvements in arm function, hand coordination, functional mobility, balance and gait function, postural control, upper-limbs function) and physical activity (Lopes, et al., 2018). Additionally, numerous rehabilitation technology companies offer game-based solutions for rehabilitation for stroke (RecoVR, Flint Rehabilitation), cognitive development (Lumosity, Intendu, NeuronUp), executive function (The Learning Corp/Constant Therapy), gross motor training and math learning for children with developmental disabilities (Zyrobotics), and more.

To be sure, electronic games are not uniformly positive in their effects on players. Numerous studies have investigated "internet gaming disorder", a disorder defined in the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders (DSM–5) and characterized by "persistent and recurrent use of the Internet to engage in games, often with other players, leading to clinically significant impairment or distress" (Feng, et al., 2017; van den Eijnden, et al., 2018). Overuse of gaming, social media and other electronic technology can lead to serious deleterious psychosocial effects. Still, positive developmental, social, intellectual, and emotional effects are also recognized by researchers. These effects can result from games other than those considered "serious games", which are commonly understood as 'any piece of software that merges a non-entertaining purpose (seriousness) with a video game structure (game)" (Djaouiti, et al., n.d.). Electronic gaming's positive effects make it critical that people with disabilities have access to these products and services. Furthermore, gaming and gaming platforms have grown in sophistication to include multi-party text, video and voice chat, which enhances the social participation opportunities for players.

To date, little research has been conducted on game playing and spectating by people with disabilities. Researchers at the University of York in the United Kingdom and AbleGamers Charity in the United States surveyed participants recruited mainly from their respective network (Beeston, et al., 2018). They found that among the 230 participants in their sample, most identified themselves as gamers (150) and considered gaming to be their primary hobby (138).

About equal numbers self-identified as hardcore gamers and casual gamers. Single-player games were the type most frequently played, but multiplayer online games were also frequently played. Only 33 participants indicated that they did not play any online multiplayer games compared to 82 who said they did not play local (co-located) multiplayer games. Respondents indicated they play for multiple reasons, including having fun, to relax, be part of a community and other reasons.

Methods

Survey data were collected from August 27 to October 6, 2020 using convenience sampling. Participants were recruited via the Consumer Advisory Network (CAN) and the Accessibility User Research Collective (AURC), both maintained by researchers in the Rehabilitation Engineering Research Center for Community Living, Health and Function (LiveWell RERC) at Shepherd Center, a rehabilitation hospital in Atlanta, Georgia. All members of these U.S.-based national networks of people with disabilities are 18 years or older and report having a disability or functional limitation. It was not necessary to play electronic games to participate in the survey, but the substantive focus of the study may have skewed sampling in favor of people who play electronic games.

All participants completed the questionnaire online, even though our recruitment messages invited participants to call the study team if they wished to have assistance via voice communication. Survey Monkey served as the data collection platform. The study team has previously tested the screen-reader accessibility of Survey Monkey and found that most types of questions are screen-readers accessible, with the main exception being “grid” style questions asking respondents to rate multiple options within the same question. This style of question was not included in the questionnaire.

The questionnaire has several sections:

- Demographics, disability, and technology profile
- Gaming activities
- eSports viewing
- Video games and rehabilitation
- Social interactions

Demographic and disability profile and overall participation in electronic gaming for the sample are presented in Table 1. While demographic variables are not used in our statistical analyses for this paper, the statistics are included to provide an overview of study representation.

Table 1. Demographic and Study Variable Summary Statistics (N=402)

Demographics	**Statistics**
Mean age in years (standard deviation)	46.59 (sd=14.91)
Sex (Female %)	49.9%
Marital status (Married %)	41.1%
Race (Non-White %)	29.9%
Education (minimum Bachelors degree)	65.5%
Plays Video Games	58.8%

Data: Survey on Video Gaming, eGaming and eSports, 2020.

The disability types represented in our sample are summarized in Table 2. The questionnaire developed for this survey research project included 17 distinct functional limitations. However, it was impractical to provide analysis for all these limitations in this article. Instead, we focused on disabilities most related to ICT access. Most people reported having more than one disability or functional limitation. The questionnaire was designed to allowed multiple responses to questions on self-identified disabilities. This is reflected in the percentages in Table 2, which add to 155.6%.

Table 2. Disability Types of Sample (n= 402)

Disability Type	Percent of sample
Learning Disability	18.5%
Anxiety	22.6%
Speaking Limitation	6.6%
Upper Extremity Limitation	22.0%
Walking Limitation	29.2%
Fatigue & Limited Stamina	23.1%
Blind	29.2%
Deaf	4.4%

Source: Survey on Video Gaming, eGaming and eSports, 2020.

Analysis

The relationship between disability type and video game playing was tested using phi, (Φ; Fleiss and Berlin, 2009) to understand if respondents with a specific type of disability played video games at a higher percent than all other respondents with disabilities (Table 3). Respondents with learning disabilities reported playing video games at high levels (81%) and were significantly more likely to play video games than respondents who did not have a learning disability (Φ= 0.225, p<.001). Statistically significant patterns also exist for respondents with anxiety (72.8%, Φ= 0.162, p<.01); and limitations with fatigue and limited stamina (69.5%, Φ= 0.126, p<.05). People who are blind were significantly less likely to play video games (46.9%, Φ= -0.161, p<.01). For all other disability types reported in Table 3, there was not a significant difference in video game use, either marginally or with traditional cut-off values.

Table 3. Electronic Game Playing and Disability Type

Disability Type	Disability?	Plays video games – No	Plays video games – Yes	N	Φ	Significance
Learning Disability	No Yes	46.7% 19.0%	53.3% 81.0%	338 84	.225	p<.001

Disability Type	**Disability?**	**Plays video games – No**	**Plays video games – Yes**	**N**	**Φ**	**Significance**
Anxiety	No Yes	45.8% 27.2%	54.2% 72.8%	319 103	.162	p<.01
Speaking Limitation	No Yes	41.8% 33.3%	58.2% 66.7%	392 30	.044	*
Upper Extremity Limitation	No Yes	42.9% 36.0%	57.1% 64.0%	322 100	.059	*
Walking Limitation	No Yes	41.2% 41.4%	58.8% 58.6%	289 133	-.002	*
Fatigue & Limited Stamina	No Yes	44.8% 30.5%	55.2% 69.5%	317 105	.126	p<.05
Blind	No Yes	36.0% 53.1%	64.0% 46.9%	292 130	-.161	p<.01
Deaf	No Yes	41.3% 40.0%	58.7% 60.0%	402 20	.909	*

*p≥0.10

Source: Survey on Video Gaming, eGaming and eSports, 2020.

Respondents were asked to identify their favorite type of electronic game (Table 4). Each cell reports the percent of respondents with specific disability for whom each of the 12 electronic game types is their favorite. Overall, the favorite type of game by all respondents with any type of disability is word games (26.7%). There are notable differences between disability types. For example, blind participants identified “other games” (those not listed among the responses) as their favorite type of game (28.1%), followed by word games (24.6%), and role-playing and action games (each by 8.8% of blind respondents). In contrast, people who are deaf preferred role-playing games at the same rate as their preference for word games (25%), the highest rate for this group. Deaf respondents reported that their third favorite game type was action games (16.7%). A substantial percentage of respondents with a learning disability or anxiety also identified role playing games as their preferred game type (13.6% and 16.4%, respectively) after word games and “other” games. Action and adventure games receive the most attention in commercial advertisements, especially on television. Notably, respondents with speaking limitations were the most likely to identify these games as their favorite (25% for both types combined), followed by those with a learning disability (21.2%), those who are deaf (16.7%), those with anxiety (15%) and those with fatigue and limited stamina (14.5%).

The least favorite type of games are exergames (0.4%) and virtual reality (0.7%). For exergames, only a small percentage of people who are blind (3.5%) identified exergames as their favorite. Exergames were not identified by any other disability type as their favorite. For virtual reality games a small percentage of respondents who have walking limitations (2.7%), learning disabilities (1.5%), and fatigue and limited stamina (1.4%) identified these as their favorite types of games. The low popularity of VR games might be a result of the higher cost for additional devices, accessories, and the games themselves.

We were also interested in the social aspects of gaming. Table 5 shows with whom respondents play electronic games, by disability type. Respondents were asked to identify each group of people with who they play. Rows do not sum to 100.0%, as respondents were asked to identify all that apply. Overall, respondents play electronic games by themselves most frequently (38.3%). Those with a learning disability are most likely to play alone (47.6%), followed by those with upper extremity limitation (45%), and those with fatigue and limited stamina (44.8%).

When playing games with others, respondents overall play most frequently with people they met online (27.1%) or with friends (27.0%). Playing with friends was most common for people who report having anxiety (36.9%), deafness (35.0%), or a learning disability (34.5%). Notably, people with either a speaking limitation or a physical/mobility limitation report playing with people they met online more frequently than with friends. Playing with acquaintances met online is most common for respondents with a speaking limitation (33.3%) and respondents who are deaf (30.0%), two functional limitations which might be related.

Table 4. Favorite Electronic Game Types and Disability Type

Disability Type	Action	Adventure	Battle Royale	Racing	Role Playing	Simulation	Strategy	Sports	Online massive multiplayer	Exergame	Virtual Reality	Word Games	Other	N
Learning Disability	10.6%	10.6%	1.5%	7.6%	13.6%	1.5%	3.0%	3.0%	6.1%	0.0%	1.5%	25.8%	15.2%	66
Anxiety	8.2%	6.8%	4.1%	4.1%	16.4%	1.4%	5.5%	2.7%	4.1%	0.0%	0.0%	28.8%	17.8%	73
Speaking Limitation	15.0%	10.0%	0.0%	0.0%	5.0%	0.0%	5.0%	0.0%	5.0%	0.0%	0.0%	30.0%	30.0%	20
Upper Extremity Limitation	8.1%	4.8%	0.0%	8.1%	16.1%	1.6%	9.7%	1.6%	8.1%	0.0%	0.0%	21.0%	21.0%	62
Walking Limitation	6.7%	5.3%	0.0%	6.7%	9.3%	1.3%	6.7%	5.3%	9.3%	0.0%	2.7%	28.0%	18.7%	75
Fatigue & Limited Stamina	8.7%	5.8%	0.0%	5.8%	10.1%	1.4%	2.9%	1.4%	8.7%	0.0%	1.4%	30.4%	23.2%	69
Blind	8.8%	1.8%	1.8%	1.8%	8.8%	1.8%	5.3%	7.0%	7.0%	3.5%	0.0%	24.6%	28.1%	57
Deaf	16.7%	0.0%	0.0%	0.0%	25.0%	0.0%	0.0%	0.0%	16.7%	0.0%	0.0%	25.0%	16.7%	12
Average	10.4%	5.6%	0.9%	4.3%	13.0%	1.1%	4.8%	2.6%	8.1%	0.4%	0.7%	26.7%	21.3%	--

Source: Survey on Video Gaming, eGaming and eSports, 2020.

Playing with family members (13.9%) and video game leagues (2.1%) are the least common types gaming partners/scenarios. Playing with family is most frequently reported by respondents with anxiety (21.4%), a speaking limitation (20.0%) or fatigue limited stamina (20.0%). Respondents with speaking limitations are the most active players in leagues (6.7%), followed by those who are deaf (5.0%), have upper extremity limitation (5.0%) or have fatigue/limited stamina (4.8%). As league play grows in popularity and becomes more accessible, the rate of participation by people with disabilities might grow.

Table 5. Electronic Gaming Partners and Disability Type (%)

Disability Type	Plays Alone	With Friends	Family	People Met Online	League Play	Other
Learning Disability	47.6%	34.5%	17.9%	28.6%	1.2%	3.6%
Anxiety	40.8%	36.9%	21.4%	28.2%	2.9%	2.9%
Speaking Limitation	40.0%	20.0%	20.0%	33.3%	0.0%	6.7%
Upper Extremity Limitation	45.0%	21.0%	13.0%	27.0%	2.0%	5.0%

Disability Type	Plays Alone	With Friends	Family	People Met Online	League Play	Other
Walking Limitation	38.3%	21.8%	15.0%	22.6%	2.3%	3.0%
Fatigue/Limited Stamina	44.8%	22.9%	20.0%	27.6%	3.8%	4.8%
Blind	27.1%	24.1%	12.0%	19.5%	1.5%	0.8%
Deaf	25.0%	35.0%	15.0%	30.0%	5.0%	5.0%
Average	38.6%	27.0%	16.8%	27.1%	2.3%	4.0%

Source: Survey on Video Gaming, eGaming and eSports, 2020.

Finally, knowing why people with disabilities play electronic games is critical to understanding motivation and behaviors. Table 6 shows responses for all participants, who were asked to select all reasons for playing that apply. An overwhelming majority (91.5%) said they play electronic games for entertainment or fun. Approximately three-fifths of respondents said they play to take a break from reality (62.1%) or to reduce stress (59.7%). Slightly less than half of the respondents said they play for social activity (44.0%).

Notably, few respondents said they play for training/skill building (11.7%), physical activity/exercise (10.5%), education (8.9%) or rehabilitation (6.5%). These results are consistent with the low numbers of respondents who report playing exercise games. It also suggests that rehabilitation professionals (clinical, engineering and research) might expect substantial challenges engaging people with disabilities in games by emphasizing functional improvement or training. This challenge might be especially difficult for games that are specifically designed for rehabilitation and recovery/maintenance of function. Our survey results instead point to the inherently therapeutic and immersive nature of electronic games versus games designed specifically as therapy. Instead of “gamification” of therapy, it might be more effective to consider gaming as therapy.

Table 6. Reason for Playing Electronic Games

Reason	**N**	**%**
Entertainment or fun	227	91.5%
Take a break from reality	154	62.1%
Reduce stress	148	59.7%
Social activity	109	44.0%
Art/aesthetic experience	40	16.1%
Competitive experience	32	12.9%
Training and simulated skill building	29	11.7%
Physical activity/exercise	26	10.5%
Education	22	8.9%
Rehabilitation or tele-rehab	16	6.5%
Other	16	6.5%

Source: Survey on Video Gaming, eGaming and eSports, 2020.

Conclusion

This exploratory research study is the first step in mapping the electronic gaming terrain for people with disabilities. Our data show that a substantial number of people with disabilities participate in e-gaming. But overall, they play electronics games mostly alone, they do not play primarily for social engagement and do not play to enhance or maintain functional performance or ability. In the end, people with disabilities just play to have fun, take a break from reality and reduce stress. Our survey data also show that some variation in the gaming experiences and preferences among people with specific disabilities do exist. Ongoing inquiry and analysis of our survey results will provide additional insights going forward.

Works cited

Beeston, Jen, Power, Christopher, Cairns, Paul, and Barlet, Mark. “Characteristics and Motivations of Players with Disabilities in Digital Games.” (Unpublished manuscript), 2018, https://arxiv.org/abs/1805.11352, Accessed 05 Sep 2020.

Consumer Technology Association. “2019 Future of Gaming.” https://shop.cta.tech/products/2019-future-of-gaming. Accessed 06 Sep 2020.

Djaouti, Damien, Alvarez, Julian, Jessel, Jean-Pierre. “Classifying Serious Games: The G/P/S Model.” http://www.ludoscience.com/files/ressources/classifying_serious_games.pdf. Accessed 05 Sep 2020.

Feng, Wendy, Ramo, Danielle, Chan, Steven, and Bourgeois, James. “Internet Gaming Disorder: Trends in Prevalence 1998–2016.” *Addictive Behavior*, vol. 75, 2017, pp. 17–24, doi:10.1016/j.addbeh.2017.06.010

Fleiss, J. L., and Berlin, J. A. “Effect sizes for dichotomous data.” *The Handbook of Research Synthesis and Meta-Analysis*, edited by H. Cooper, L. V. Hedges, and J. C. Valentine, 2009, Russell Sage Foundation. pp. 237–253.

France, Christopher R., Thomas, James S. (2018). Virtual immersive gaming to optimize recovery (VIGOR) in low back pain: A phase II randomized controlled trial. *Contemporary Clinical Trials*, vol. 69, 2018, pp. 83–91, doi:10.1016/j.cct.2018.05.001.

Hodge, SE, Taylor, J. and McAlaney, J. “Is It Still Double Edged? Not for University Students’ Development of Moral Reasoning and Video Game Play.” *Front. Psychol.,* vol. 11, 2020, pp. 1313, doi:10.3389/fpsyg.2020.01313

IFPI. (N.D.) Global Music Report. https://www.ifpi.org/ifpi-issues-annual-global-music-report/. Accessed 05 Sep 2020.

Jones, Katie. (2020). “Online Gaming: The Rise of a Multi-Billion Dollar Industry.” Visual Capitalist. https://www.visualcapitalist.com/online-gaming-the-rise-of-a-multi-billion-dollar-industry. Accessed 04 Sep 2020.

Kowert, Rachel, Domahidi, Emese, Quandt, Thorsten. (2014) The Relationship Between Online Video Game Involvement and Gaming-Related Friendships Among Emotionally Sensitive Individuals. *Cyberpsychology, Behavior and Social Networking*, vol. 17, no. 7, 447-353, doi: 10.1089/cyber.2013.0656

Lopes, Sílvia, Magalhães, Paula, Pereira, Armanda, Martins, Juliana, Magalhães, Carla, Chaleta, Elisa and Rosário, Pedro. “Games Used with Serious Purposes: A Systematic Review of Interventions in Patients with Cerebral Palsy.” *Front. Psychol.*, vol. 9, 2018, https://doi.org/10.3389/fpsyg.2018.01712

Statista. “Consumer spending on gaming in the United States from 2010 to 2018, by segment.” https://www.statista.com/statistics/190195/total-consumer-spend-on-us-games-industry-2010/. Accessed 05 Sep 2020.

Statista. “Video Streaming (SVoD).” Online: https://www.statista.com/outlook/206/100/video-streaming--svod-/worldwide. Accessed 04 Sep 2020.

van den Eijnden, Regina, Koning, Ina, Doornwaard, Suzan, van Gurp, Femke, Ter Bogt, Tom. (2018). The impact of heavy and disordered use of games and social media on adolescents’ psychological, social, and school functioning. *Journal of Behavioral Addictions* vol. 7, no. 3, 2018, pp. 697–706, doi:10.1556/2006.7.2018.65

2021
The Journal on Technology and Persons with Disabilities

Evaluating Cognitive Complexity of Algebraic Equations

Akashdeep Bansal, M. Balakrishnan
Indian Institute of Technology Delhi, India
mbala@cse.iitd.ac.in, akashdeep@cse.iitd.ac.in
Volker Sorge
University of Birmingham, United Kingdom
v.sorge@cs.bham.ac.uk

Abstract

While blind and visually impaired (BVI) readers can access digital text information using auditory interface through screen readers, accessing mathematical equations is much more challenging task, which is known to limit access to STEM subjects for BVI students. This is primarily due to the complexities of equation layout as well as the fact that only a small number of symbols can lead to a very lengthy speech string resulting in a high cognitive load for the reader. One approach to reduce this load can be to abstract parts of an equation/expression via some abbreviating symbol which is then read out separately. This division/partitioning into expressions is based on computing the cognitive complexity as "felt" by the reader.

In this paper we present a study that aims to identify a suitable metric for this cognitive complexity. To assess this complexity, we have conducted an extensive user study with eighteen VI persons. In our study we captured users' response in terms of errors in reproducing equations, time taken to understand an equation, etc. We used this data to propose candidate metrics for measuring this complexity. We correlated these metrics with complexity measures available in open-source systems for math accessibility. We have been able to identify a suitable complexity metric that correlates well with results obtained from user study and thus can be integrated to screen readers while accessing equations.

Keywords

Cognitive Complexity, Mathematical Equations, Visually Impaired, Auditory User Interface

Introduction

Providing access to mathematical equations for blind or visually impaired (BVI) students is essential for inclusive STEM education. While access to text, even scientific writing, can be achieved quite straightforwardly using regular screen reading software, for mathematical equations this is more complicated. While mathematics can be voiced by a number of screen readers and specialised tools, the non-linearity of formulas using two dimensional layout with sub- and superscripts, fraction, square-root, parenthesis, etc. means that faithfully and unambiguously voicing an equation leads to a very lengthy speech output. This verbosity does not only lead to a considerable cognitive load but also takes valuable reading time leaving students with visual impairments at a clear disadvantage compared to their sighted peers.

Though this work is presented in the context of BVI users, but it is clear that auditory interfaces are getting popular with non-BVI readers as well. Utilizing the time for studying when it is not possible to visually focus on text (like driving or exercising) is becoming prevalent. Thus, it is possible to use this study in the context of any situation where auditory interfaces are preferred choice for accessing equations.

Even simple equations like the quadratic formula $x = \frac{-b \pm \sqrt{b^2-4ac}}{2a}$ can, when interpreted correctly and put into speech, result in somewhere between *25* and *30* words in spoken text, depending on the system, choice of speech rules and preference settings. Clearly, understanding such an explanation is not only time consuming but also very difficult. Moreover, it is strictly not necessary to go through the complete equation for casual reading. When seeing the equation, while still complex, the reader can quite easily understand that there is a variable on the left-hand side and a fraction on the right-hand side. This is generally enough to gain an initial understanding, and in case more details are needed the equation can be examined in more depth.

In order to speed up voicing of equations while still avoiding ambiguity, various cues have been proposed in audio rendering of mathematics, ranging lexical (Chang 1983) and prosodic (Raman 1994) techniques to earcons (Stevens 1997), spearcons (Bates 2010), audio spatialization (Potluri 2014), and auditory cues (Murphy 2010), etc. While some of them also try to optimize for cognitive load, verbosity, ambiguity, and naturalness, as the complexity of equations increases, rendering with these proposed cues becomes more and more ineffective. Indeed a number of user studies (Potluri 2014, Stevens 1997, Frankel 2014, Gellenbeck 2009, Murphy 2010) have found that none of the cues in isolation or combination were able to make rendering completely unambiguous, natural, succinct, and cognitively efficient.

Consequently, alternatives have been proposed that decompose mathematical expressions into smaller, more manageable parts. Hierarchical navigation (Sorge 2014, Frankel 2014), summarization (Gillan 2004, Cervone 2016), operator abstraction (Cervone 2016) and variable substitution (Raman 1994) are some of these techniques. Hierarchical navigation allows the stepwise exploration of sub-expressions. Summarization verbally abbreviates parts of formulas. Operator abstraction and variable substitution replace entire sub-expressions and render – aurally and visually – much reduced formulas. These latter techniques are commonly employed in mathematical practice as well as in mainstream computer algebra systems such as Mathematica (Mathematica 2009), and Maple (Maple 2016), where very long equations are abbreviated for display on the screen using variable substitution. For accessibility the AsTeR system (Raman 1994) provides variable abstraction in large formulas on the basis of a metric that attaches weights to a modified LaTeX tree. The MathJax (Cervone 2016, Cervone 2016) rendering library can abstract sub-expressions visually on the basis of the rendered size of semantically defined chunks and then these can also be aurally summarized.

Common to all these approaches is that the chosen abstraction is based on complexity metrics that are computed with respect to the internal representation of a formula in that particular system. While these can range from purely syntactic to partial semantic interpretations, it is unclear that they in any way coincide with the cognitive load experienced by a reader. In this paper we present an analysis that aims to shed some light on this relationship.

In particular, we present a user study where we worked with eighteen BVI students and professionals to capture their process of understanding of mathematical equations when accessing them with a screen reader. Readers were allowed multiple attempts at reproducing equations, while their responses were recorded and analyzed in terms of cumulative errors, attempts and time taken for comprehension, etc. We use this data to propose candidate metrics for measuring cognitive complexity and compare it with similar complexity metrics that are available in open source systems for mathematics accessibility. While our current study is restricted to relatively simple algebraic equations and can be seen as an initial step towards a cognitive model of understanding mathematical equations that will help in the development of adaptable accessibility technology for that domain. In particular, it aims to provide answers to the questions on how the cognitive load on the reader of an equation can be evaluated and to what extent the complexities currently used by assistive technology model this load.

The paper is structured as follows: Section “Complexity metrics” presents some of the complexity metrics that are available in the literature and existing systems and we use for comparison. Section “User study and assessment of user responses” gives an overview of the user study and an analysis of its results. We then correlate these to the metrics (from Section “Complexity metrics” in Section “Assessment of user responses” before concluding along with summary of the results and discussing some future work.

Complexity Metrics

For our study we considered the following open source systems and their respective metrics:

(1) AsTeR (Raman 1994), a system for technical reading, which is implemented in Lisp and uses LaTeX as input language.

(2) MathJax (MathJax 2020}, a JavaScript library for rendering mathematics in browsers, which uses LaTeX or MathML as input.

(3) Speech Rule Engine (SRE) (Sorge 2019), a JavaScript library for translating mathematics to speech strings, which uses MathML as input.

First, we will discuss the metrics with the example formula $\frac{-\mathrm{e}+(a^{29}+bc+d)}{\mathrm{f}}$

While the formula itself is artificially constructed, the elements it contains match those that can be found in our set of test expressions.

AsTeR computes weights for the formula in a quasi-prefix form that it derives from LaTeX expressions; the tree corresponding to our example is given in Figure 1. Note, that Aster combines operators, introduces juxtaposition operators between atoms combinations as well as combines delimited expressions. Atoms are single characters and numbers. Aster attaches a weight of one to each leaf node, regardless of its content. The weight of a node is then computed as the sum of its children plus *1*. The complexity for our example is therefore ***AsTeR****=13*.

SRE internally translates input MathML into a semantic tree, which is very similar to that of AsTeR. While the representations can differ for more complex expressions like sums, integrals, matrices, etc. As these are not of relevance for our study, we will not go into detail. One observed difference in Figure 4 is the treatment of negated atoms like *-e*. SRE allows to compute multiple complexities; it's simple one is similar to the one of AsTeR (with exception of negated atoms) and we will not consider it separately in the remainder of this work. The second, which we consider, also takes the length of atoms in leafs into account, i.e., the node *29* would have weight *2* instead of 1. Consequently, we get ***SRE***=*15* for our example.

MathJax uses complexities to determine the weight of its visual output. It thereby introduces factors to adjust for differing script sizes, for example multiplies weights for sub- and superscripts by *0.8* (see Figure 2). In addition, it considers length of atoms, giving every symbol a value of *0.5*, plus *0.5* for each leaf node. MathJax offers two ways of computing complexities: one working directly on its internal MathML-like representation (see Figure 2) and second one for an enriched version of its output, which is computed by refining its representation with SRE's semantic tree. This is given in Figure 3 and primarily used for computing visual abstractions that can be triggered in its accessibility mode. For our example, MathJax's direct complexity results in ***MJX*** =*26.16* and for the semantically enriched version in ***MJXe***=*30.96*.

We are aware that systems like MathPlayer (Soiffer 2007) also use internal weight measures, for example, to compute pause length. However, since it is not an open source system, we were unable to experiment with its metrics.

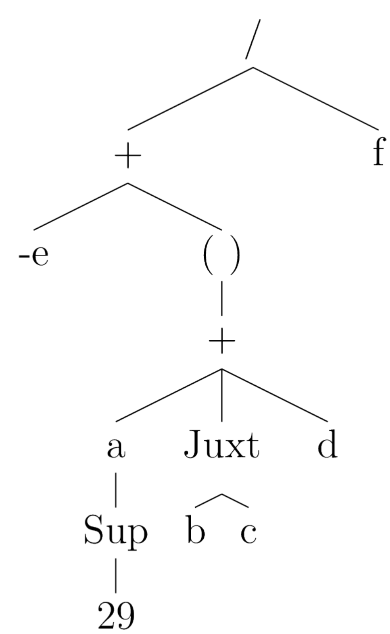

Fig. 1. Tree Representation of Aster's Quasi-Prefix Form.

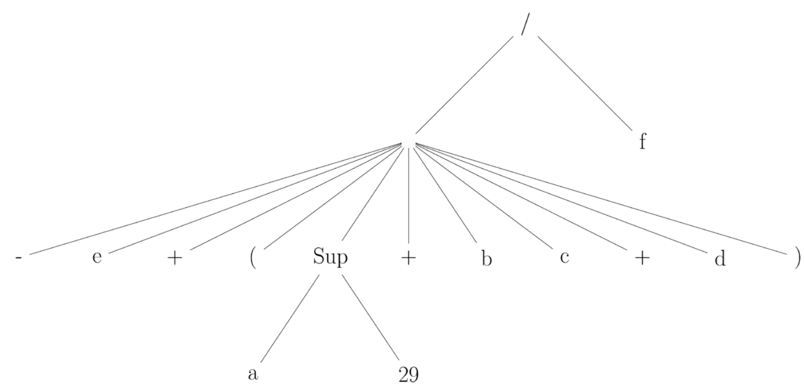

Fig. 2. MathJax's Visual Tree Representation.

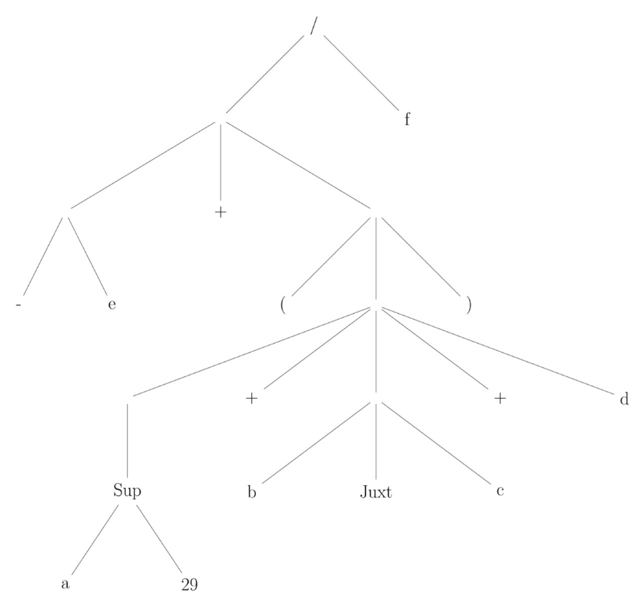

Fig. 3. MathJax's Semantically Enriched Tree.

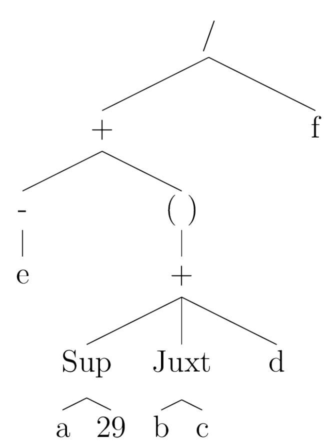

Fig. 4. SRE's Semantic Tree Representation.

User Study

The research question we pose is the following:

"Which of the proposed open source complexity metrics have highest correlation factor with the user's cognitive complexity for algebraic equations when rendered through audio?"

We have conducted a user study with eighteen BVI participants with STEM background. Participants were contacted through our network channels and accessibility mailing lists. The two main screening criteria where they should have STEM background (at least well familiar with basic algebraic equations) and should be regular screen reader users. Out of eighteen participants, thirteen were blind (some congenitally blind and some late blind) and five were with very low vision. Their ages varied from *16* to *41* years with an average age of 24.06 years and standard deviation of 5.254. The detailed demographic details of participants are given in table 1.

Table 1. The Demographics of the Participants.

ID	Age	Sex	VI	Years of VI	Occupation	Course	Year Course Completed	Screen Reader	Mode of Reading Equations	Years of Use
1	18	M	Blind	7	Student	Senior Secondary	2020	NVDA	Simplified LaTeX	1
2	21	M	Low Vision	10	Student	Under Grad Computer Science	2021	JAWS	JAWS	2
3	25	M	Low Vision	17	Student	Under Grad Information Tech	2019	JAWS	MathPlayer	4
4	26	F	Blind	Birth	Professional	Post Grad Digital Marketing	2017	JAWS	Braille	Birth
5	22	M	Blind	6	Student	Under Grad Computer Science	2021	JAWS	MathPlayer	1
6	23	M	Blind	Birth	Professional	Post Grad Math	2018	JAWS	LaTeX Source	4
7	41	M	Blind	Birth	Professional	MBA	2002	NVDA	MathPlayer	Rare

ID	Age	Sex	VI	Years of VI	Occupation	Course	Year Course Completed	Screen Reader	Mode of Reading Equations	Years of Use
8	27	M	Low Vision	Birth	Student	PhD Human Computer Interaction	NA	NVDA	MathPlayer	2
9	26	M	Blind	Birth	Professional	Post Grad Computer Science	2017	JAWS	MathPlayer	6
10	19	M	Low Vision	7	Student	Under Grad Computer Science	2023	JAWS	JAWS	1
11	28	F	Blind	14	Student	Post Grad Computer Science	2019	JAWS	Simplified LaTeX	5
12	16	M	Blind	Birth	Student	High School	2020	NVDA	MathPlayer	3
13	24	M	Blind	8	Student	Under Grad Engineering	2020	NVDA	ChattyInfty	3
14	23	M	Low Vision	6	Student	Under Grad Eco	2020	JAWS	JAWS	4
15	19	M	Blind	10	Student	Senior Secondary	2019	JAWS	JAWS	6
16	24	M	Blind	12	Student	Post Grad Math	2020	JAWS	JAWS	6
17	24	M	Blind	16	Professional	Under Grad Computer Science	2018	JAWS	MathPlayer	3
18	27	M	Low Vision	NA	Student	PhD Human Computer Interaction	NA	NVDA	MathPlayer	2

In the setup of the study we tried to eliminate bias as much as possible. Equations were artificially constructed to avoid participants being familiar with a formula from its context and prior background. Likewise, any patterns among operators and variables were avoided. All formulas used for testing were algebraic equations using only the four basic arithmetic operations and exponentiation.

30 equations of varied complexity were chosen for the test. All *30* equations are given in the first column of Table 7. Complexity was determined by a simple count of symbols and tree depth of the corresponding MathML expression. The equations were split into three equal sets of ten making sure that the variation in the various equation parameters were uniform across the sets. Six more sets were constructed utilizing the first three sets. Fourth set was formed by picking odd placed equations of set one and even placed equations of set two. Set five was constructed by picking even placed equations of set one and odd placed equations of set two. Similarly, set six and seven were created using set one and three, and set eight and nine were created using set two and three. This way, we made sure that every set has exactly ten equations and every equation appeared in exactly three sets as well as variation in the various equation parameters were uniform across the sets.

Each set was presented to two of the 18 participants using NVDA (NVAccess 2020) and MathPlayer (Soiffer 2007, Frankel 2014) with ClearSpeak (Frankel 2016) rules for voicing, ensuring that none of the systems whose metrics we wanted to test were involved. Users were asked to listen to the entire expression, without noting anything down. They would then be asked to reproduce the expression. They could repeat this procedure up to five times per equation in case they failed to correctly reproduce the equation or lacked confidence. After every attempt, they were also expected to report on their confidence about correctness assuming they were under exam conditions. Sessions were video recorded for later analysis. All participants also had gone through a training session where they were informed of the test procedure and could practice with MathPlayer in ClearSpeak settings. Overall, the study was conducted in two phases:

Phase 1 was conducted through Skype. The participants were only expected to listen and respond. The equations were audio rendered on our device. For them, the audio of our system was live-streamed. The users who participated in phase 1 are odd-numbered participants in Table 1.

Phase 2 was conducted remotely. Due to universities and corporations being shut down due to the COVID-19 pandemic lock-down, participants had migrated to their hometown, with many having poor internet connectivity. So, it was not possible to continue the study using the same protocol as phase 1. To avoid latency and disturbance due to poor internet connectivity, a web application was developed. Here, prerecorded audio clips of equations were used instead of live-streaming equation's audio from our system. An edit box was also provided to users for answering with their understanding, which gets enabled only after the audio clip ends. The value of the edit field was also set to null on requesting another utterance. We were also interacting with the user after every attempt to understand his/her understanding of the equation. The video recording of the study was still done using Skype. The web application used can be accessed at (http://iitd.info/math). The users who participated in phase 2 are even-numbered participants in Table 1.

For assessing the responses of the participants of the case study, we adapt and extend an approach presented in (Kacorri 2014) in the context of EAR-Math. It aims to access the effectiveness of the developed system through a user study by comparing the developed system's performance with the old systems or manually rendered equations by special instructors. Similar to our analysis (Kacorri 2014) employs criteria based on the errors in structure, arithmetic operations, and numbers. However, our approach is more comprehensive by also taking into account a user's confidence, number of attempts needed, and confusion of readers due to imprecision of the ClearSpeak rules.

Assessment of User Responses

We analyzed replies of participants to classify their understanding of the equations. We thereby aim to qualify the users’ ability to comprehend an equation and their ability to reproduce an equation faithfully. For this we define the following quantifiable criteria.

Structural Correctness A user is considered to have understood the structure of an equation, if position of operators, number of terms and layout features (superscript, fraction, etc.) are recalled correctly. That means individual values of terms or operators do not necessarily have to be correct. Table 2 contains some examples of what is and what is not considered structurally correct.

Table 2. Examples of Correct/Incorrect Structure.

(s means a term, which may be a variable, a variable with coefficient, or an implicit multiplication of any number of variables and constant)

Original Equation	User's Answer	Correctness	Explanation
$-o+j+y-r=s$	$s+s+s+s=s$	Yes	Layout, number of terms and operators are correct.
$-o+j+y-r=s$	$-o+j+y-r+s$	No	= is a relational operator. Its replacement has changed the structure of the equation.
$52j\frac{-o+j}{y-r}=s^p$	$s*\frac{s+s}{s}=s$	No	One term in the denominator and exponent on the right-hand side is missing.
$52j\frac{-o+j}{y-r}=s^p$	$s*\frac{s+s}{s+s}=s^s$	Yes	Layout, number of terms and operators are correct.

Reproduction Correctness As user has correctly reproduced an equation if it is structurally correct and all its terms and operators are correct as well.

Types of Errors In addition to this the errors made by users were analyzed and categorized as listed in Table 3.

Table 3. Different Types of Users Mistakes with Examples.

Error Type	Incorrect	Correct	Explanation
Primitive error	$-o+j+y-r=f$	$-o+j+y-r=s$	Term got changed
Primitive error	$3124/za=f$	$3241/za=f$	Term got changed
Relational error	$-o+j+y-r=f$	$-o+j+y-r+f$	= defines a relation, it's replacement with + has changed the relation

Error Type	Incorrect	Correct	Explanation
Operator error	$-o + j + y - r = f$	$-o + j - y - r = f$	$+$ got replaced with $-$
Structural error	$x^a + b + c$	$x^{a+b} + c$	The identification of exponent boundary is wrong but still it's layout will remain same on abstraction
Layout error	$x^a + b + c$	x^{a+b+c}	The identification of exponent boundary is wrong as well as it's layout will also change on abstraction
Permutation error	$\frac{23c + 49j}{x + 3y}$	$\frac{23c + 49j}{3x + y}$	Terms or coefficients got swapped (coefficient of x got swapped with y)
Fraction error	$x = y/z + c$	$x = y/(z + c)$	The layout of the equation got changed due to the error in identification of the fraction boundary

In addition, we found an error that was commonly committed by most users, that we finally tracked to as a problem with the ClearSpeak rule set. ClearSpeak uses three different conventions for the different types of fraction (as shown in Table 4) and utilizes pauses for boundary identification. Although this was mentioned in users training material it still posed a significant challenge for the participants. In the equations, where these errors occurred multiple times (listed in Table 5), it was only counted once against structural comprehension.

Table 4. ClearSpeak Conventions for Different Types of Fraction.

Equation	Speech String
$\frac{a + b}{c + d}$	The fraction with numerator a plus b and denominator c plus d
$a + \frac{b}{c} + d$	a plus b over c plus d
$a + b/c + d$	a plus b divided by c plus d

Table 5: Most Common Structure Comprehension Errors.

Correct Equation	**Incorrect Users' Response**
$52j\frac{567t}{sd} = 3124/149j * p$	$52j\frac{567t}{sd} = (3124 * p)/149j$
$14^p/a - 119 = z/p$	$14^p/(a - 119) = z/p$
$\frac{619t}{3124za} = 847p$	$\frac{619t}{3124z}a = 847p$
$\frac{(3a + 7z - aj)(v/b + sb - nj)}{34b/27 - az + 14f}$	$\frac{(3a + 7z - aj)(v/(b + sb - nj))}{34b/(27 - az + 14f)}$
$\frac{htq}{(49z - 12t/g)(256j + 19iz)}$	$\frac{htq}{(49z - 12t/g)}(256j + 19iz)$
$119p + 14t + \frac{343d}{12b} = um/(q + 14)$	$\frac{119p + 14t + 343d}{12b} = um/(q + 14)$

During the study, users were not informed about any of the above criteria to prevent them to focus specifically on structural comprehension or avoiding particular errors. In addition to the analysis of the reproduced equations we asked questions after every attempt to rule out typographical errors, etc.

We now define our assessment criteria on the basis of four values from the user study.

(1) Minimum number of attempts for structural comprehension

(2) Minimum number of attempts to reproduce the equation

(3) Total number of attempts taken for an equation for confident reproduction

(4) Error types and their numbers

Note, that numbers for (2) and (3) can differ, as a user can take more attempts, for confident reproduction even after correctly reproducing the equation once.

On the basis of values (1) to (4), equations are ranked for each individual user following two criteria:

- **C1: Ability to comprehend:** This gives preference order to criteria (1), over (2), (3), (4).
- **C2: Ability to reproduce:** This gives preference order to (2) over (1), (3), (4).

In other words, the difference between C1 and C2 is the preference of structural comprehension and faithful reproduction, respectively. Table 6 shows an illustration of ranking equations based on criteria C1 and C2. For example, for equation *2*, the user took *2* attempts, whereas for equation *3*, the same user took *5* attempts. Thus under criterion C1, Equation *3* ranks lower than *2*, whereas it ranks higher under criterion C2. Wherever, we could not distinguish equations amongst multiple ranks, we assigned the average value for the ranks to each equation. For example, if two equations could not be distinguished in rank *2* and *3*, we assigned *2.5* to both equations.

The average ranking of equations based on the criteria C1 and C2 are shown in Table 7 and 8, respectively. Thereby the assigned rankings were averaged across all the users, which are considered as user complexities as per criteria C1 and C2, respectively. This average should help reducing any bias which may occur due to user's prior experiences, educational background, etc., and also the variation among the different sets.

Table 6. Example for Ranking Criteria.

Eq No	**No. of attempts for Structure**	**No. of attempts to Reproduce**	**Total no. of attempts**	**Rank in C1**	**Rank in C2**
1	1	2	3	1	1
2	2	2	2	3	2
3	1	5	5	2	3

Table 7. Average Participant Ranking of Equations Based on Criterion C1.

Equation	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Avg
$21m/343g = q$	1.5	1					2	1			1.5	1.5							1.42
$823^{x+g} = z$	1.5	4.5					1	2.5			1.5	1.5							2.08
$4^a + 29^q = 20$	3	2.5							1.5	1			3	2					2.17
$p + z - x - a = w$			1	1					1.5	2.5					7	1.5			2.42
$vl = \frac{23g}{124bt}$			3	3			3	4									3	1	2.83
$619t/3124za = 847p$					3	8							1	1	3	1.5			2.92
$-e - x + j = c - z$					6	1							5	6			1	4	3.83
$\frac{246j * 17h}{dlr} = t$	6	2.5							7	2.5			2	3					3.83
$(6h - f)/v = 14 + d$	6	4.5					5	2.5			3.5	4							4.25
$496u/sd = 27g * bc$			8	5.5					5	5							2	2	4.58
$1123^{l-z} = \frac{123}{t + j}$			4	4					6	5					4	5			4.67
$-o + j + y - r = f$	4	6.5					4	5			3.5	5							4.67
$52j\frac{567t}{sd} = 3124/149j * p$					1	5					10	3			2	7.5			4.75
$k + c - z - p = q - a + m$	6	6.5							3.5	5			4	4					4.83
$14^p/a - 119 = z/p$					4	4					9	9			1	4			5.17
$14^p - 168/t = 243^r$			2	5.5			9	8									4	5	5.58
$-a + q + t - p = -l - d - k$			7	2			7.5	6							8	3			5.58
$d - 31 = (57w + 88y)bz$			5	8					3.5	7					5	6			5.75

Equation	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Avg
$6731^{\frac{z}{j}} = \frac{87x+o}{zp}$					5	6							7	9			6	3	6
$\frac{87u+74p+18m}{36d-20p-199w} = or + \frac{n}{31}$					2	2.5					6	10			9	7.5			6.17
$(714t-97q)/(13j+49x) = 26l/481$					8	2.5							6	5			8	9	6.42
$u+k-e-q = l+c+y-g$					7	7					7	6					5	8	6.67
$\frac{(24a+3q)(4z+343t)}{23c(hr+45)}$	9	8					6	7			5	7.5							7.08
$670 - lq = \frac{kp+14r-119h}{2mf-32c}$			6	7			7.5	9.5									9	6	7.5
$\frac{47-5/m}{128^{z/b}} = \frac{n}{18} + \frac{l}{s} - \frac{14}{vb}$					9	9					8	7.5			6	9			8.08
$(12j - \frac{45}{v})^{x+32/t} = \frac{n-234}{xas}$			9	10			10	9.5									7	7	8.75
$119p + 14t + \frac{343d}{12b} = um/(q+14)$	10	9							9	8			10	8					9
$\frac{3+\frac{31}{p}}{24xg} = (3b-179u)^{\frac{q}{t}}$	8	10							8	9			9	10					9
$\frac{(3a-7z+aj)(\frac{v}{b}+sb-nj)}{\frac{34p}{27}-az+14f}$					10	10							8	7			10	10	9.17
$\frac{htq}{(49z-12t/g)(256j+119iz)}$			10	9					10	10					10	10			9.83

Table 8. Average Participant Ranking of Equations Based on Criterion C2.

Equation	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Avg
$21m/343g = q$	1.5	1					4	1			1.5	1.5							1.75
$p + z - x - a = w$			1	1					1.5	2.5					2	3.5			1.92
$4^a + 29^q = 20$	3	2.5							1.5	1			2	2					2
$823^{x+g} = z$	1.5	4.5					1	2.5			1.5	1.5							2.08
$vl = \frac{23g}{124bt}$			3	3			5	4									3	1	3.17
$-e - x + j = c - z$					5	3							4	5			1	4	3.67
$\frac{246j * 17h}{dlr} = t$	6	2.5							7	2.5			1	3					3.67
$14^p/a - 119 = z/p$					3	5					8	4			1	1			3.67
$(6h - f)/v = 14 + d$	6	4.5					3	2.5			3.5	3							3.75
$619t/3124za = 847p$					2	8							5	1	4	3.5			3.92
$1123^{l-z} = \frac{123}{t + j}$			4	4					6	5					5	2			4.33
$-o + j + y - r = f$	4	6.5					2	5			3.5	5							4.33
$496u/sd = 27g * bc$			7	5.5					5	5							2	2	4.42
$k + c - z - p = q - a + m$	6	6.5							3.5	5			3	4					4.67
$6731^{\frac{z}{j}} = \frac{87x + o}{zp}$					4	1							6	8			6	3	4.67
$14^p - 168/t = 243^r$			2	5.5			9	8									4	5	5.58
$u + k - e - q = l + c + y - g$					7	2					6	6					5	9	5.83
$-a + q + t - p = -l - d - k$			8	2			6.5	6							8	5			5.92
$d - 31 = (57w + 88y)bz$			5.5	8					3.5	7					6	6			6

Equation	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Avg
$52j\frac{567t}{sd} = 3124/149j * p$					1	4					10	9			3	9			6
$670 - lq = \frac{kp + 14r - 119h}{2mf - 32c}$			5.5	7			6.5	9.5									9	6	7.25
$\frac{(24a + 3q)(4z + 343t)}{23c(hr + 45)}$	9	8					8	7			5	7.5							7.42
$\frac{47 - 5/m}{128^{z/b}} = \frac{n}{18} + \frac{l}{s} - \frac{14}{vb}$					8	9					9	7.5			7	7			7.92
$(714t - 97q)/(13j + 49x) = 26l/481$					9	6.5							7	9			8	8	7.92
$\frac{87u + 74p + 18m}{36d - 20p - 199w} = or + \frac{n}{31}$					6	6.5					7	10			9	9			7.92
$(12j - \frac{45}{v})^{x+32/t} = \frac{n - 234}{xas}$			9	10			10	9.5									7	7	8.75
$119p + 14t + \frac{343d}{12b} = um/(q + 14)$	10	9							9	8			10	7					8.83
$\frac{(3a - 7z + aj)(\frac{v}{b} + sb - nj)}{\frac{34p}{27} - az + 14f}$					10	10							8	6			10	10	9
$\frac{3 + \frac{31}{p}}{24xg} = (3b - 179u)^{\frac{q}{t}}$	8	10							8	9			9	10					9
$\frac{htq}{(49z - 12t/g)(256j + 119iz)}$			10	9					10	10					10	9			9.67

We now correlate these user complexities to the complexity values discussed in section "Complexity metrics". Firstly, in addition to the user complexities based on criteria C1 and C2, we also compute their average **Avg**. This is important as it gives equal weight to structural comprehension and reproduction of the equation. Secondly, for each equation we compute the complexities ***AsTeR***, ***MJX***, ***MJXe***, and ***SRE***. All these values are given in Table 9 and 10 as absolute and normalized values, respectively.

In the next step, given in Table 11 and 12, we compute the average rank for each metric, meaning that we can now order the equations with respect to their complexity for each metric. We can observe that equations *6*, *10*, *13*, *15*, *16*, *20*, *21*, *22*, *24* have some significant differences in ranks for C1 and C2. While equations *6*, *13*, *16*, *21*, and *22* have higher rank for C2, equations *10*, *15*, *20*, and *24* have higher rank for the C1. Apart from the fact that most of these equations contain fractional expressions, there appears no specific characteristic difference that determines why rank higher in C1 or C2.

In the context of our study, the absolute value of the complexity does not matter much. What matters is their relative values vis-a-vis others. So, we also clustered equations using K-means algorithm into five categories (0-4), where 0 corresponds to a set of easy equations while four corresponds to a set of highly complex equations. Here, the algorithm was applied on normalized complexity values of the various metrics. This way, each equation got a cluster-ID corresponding to each metric. The variation in the normalized complexity values, average rank, and cluster IDs for various equations are shown in Figures 5, 6, and 7, respectively.

Table 9. Absolute Complexity Values of Equations Based on the Various Metrics.

Equation	C1	C2	Avg	*MJX*	*MJXe*	*SRE*	*AsTeR*
$21m/343g = q$	1.42	1.75	1.58	15.5	22.5	12	9
$4^a + 29^q = 20$	2.17	2	2.08	20.6	21.6	11	9
$823^{x+g} = z$	2.08	2.08	2.08	15.8	15.8	9	7
$p + z - x - a = w$	2.42	1.92	2.17	18	20	8	8

Equation	**C1**	**C2**	**Avg**	***MJX***	***MJXe***	***SRE***	***AsTeR***
$vl = \frac{23g}{124bt}$	2.83	3.17	3	19.8	27.6	15	12
$619t/3124za = 847p$	2.92	3.92	3.42	21.5	33.5	19	12
$-e - x + j = c - z$	3.83	3.67	3.75	20	23	10	9
$\frac{246j * 17h}{dlr} = t$	3.83	3.67	3.75	22.6	30.6	17	14
$(6h - f)/v = 14 + d$	4.25	3.75	4	24.5	31.5	13	12
$14^p/a - 119 = z/p$	5.17	3.67	4.42	24.3	27.3	14	11
$-o + j + y - r = f$	4.67	4.33	4.5	20	24	9	8
$1123^{l-z} = \frac{123}{t + j}$	4.67	4.33	4.5	25.3	25.3	16	11
$496u/sd = 27g * bc$	4.58	4.42	4.5	23.5	37.5	18	15
$k + c - z - p = q - a + m$	4.83	4.67	4.75	26	30	12	12
$6731^{\frac{z}{j}} = \frac{87x + o}{zp}$	6	4.67	5.33	28.26	32.26	19	15
$52j\frac{567t}{sd} = 3124/149j * p$	4.75	6	5.38	32.8	46	26	18
$14^p - 168/t = 243^r$	5.58	5.58	5.58	26.1	28.1	16	11
$-a + q + t - p = -l - d - k$	5.58	5.92	5.75	30	34	13	11
$d - 31 = (57w + 88y)bz$	5.75	6	5.88	27.5	41.5	18	15
$u + k - e - q = l + c + y - g$	6.67	5.83	6.25	30	35	13	13
$\frac{87u + 74p + 18m}{36d - 20p - 199w} = or + \frac{n}{31}$	6.17	7.92	7.04	47.6	66	37	29
$(714t - 97q)/(13j + 49x) = 26l/481$	6.42	7.92	7.17	44	65	31	23
$\frac{(24a + 3q)(4z + 343t)}{23c(hr + 45)}$	7.08	7.42	7.25	40.8	62.4	32	27
$670 - lq = \frac{kp + 14r - 119h}{2mf - 32c}$	7.5	7.25	7.38	39.6	58	32	26
$\frac{47 - 5/m}{128^{z/b}} = \frac{n}{18} + \frac{l}{s} - \frac{14}{vb}$	8.08	7.92	8	49.44	53.84	30	25
$(12j - \frac{45}{v})^{x+32/t} = \frac{n - 234}{xas}$	8.75	8.75	8.75	47.5	55.5	28	23
$119p + 14t + \frac{343d}{12b} = um/(q + 14)$	9	8.83	8.92	44.2	60.4	30	23
$\frac{3 + \frac{31}{p}}{24xg} = (3b - 179u)^{\frac{q}{t}}$	9	9	9	44.44	54.64	27	23
$\frac{(3a - 7z + aj)(\frac{v}{b} + sb - nj)}{\frac{34p}{27} - az + 14f}$	9.17	9	9.08	54.88	80.16	42	39
$\frac{htq}{(49z - 12t/g)(256j + 119iz)}$	9.83	9.67	9.75	38	58	31	25

Table 10. Normalized Complexity Values of Equations Based on the Various Metrics.

Equation	**C1**	**C2**	**Avg**	***MJX***	***MJXe***	***SRE***	***AsTeR***
$21m/343g = q$	-1.79	-1.6	-1.71	-1.37	-1.05	-0.89	-0.97
$4^a + 29^q = 20$	-1.46	-1.49	-1.49	-0.91	-1.1	-0.99	-0.97
$823^{x+g} = z$	-1.5	-1.45	-1.49	-1.34	-1.44	-1.21	-1.22
$p + z - x - a = w$	-1.35	-1.52	-1.45	-1.14	-1.19	-1.32	-1.09
$vl = \frac{23g}{124bt}$	-1.17	-0.99	-1.09	-0.98	-0.74	-0.57	-0.58
$619t/3124za = 847p$	-1.13	-0.67	-0.91	-0.83	-0.39	-0.14	-0.58
$-e - x + j = c - z$	-0.73	-0.78	-0.76	-0.96	-1.02	-1.1	-0.97
$\frac{246j * 17h}{dlr} = t$	-0.73	-0.78	-0.76	-0.73	-0.56	-0.35	-0.32
$(6h - f)/v = 14 + d$	-0.55	-0.74	-0.65	-0.56	-0.51	-0.78	-0.58
$14^p/a - 119 = z/p$	-0.15	-0.78	-0.47	-0.58	-0.76	-0.67	-0.71
$-o + j + y - r = f$	-0.36	-0.5	-0.44	-0.96	-0.96	-1.21	-1.09
$1123^{l-z} = \frac{123}{t + j}$	-0.36	-0.5	-0.44	-0.49	-0.88	-0.46	-0.71
$496u/sd = 27g * bc$	-0.4	-0.46	-0.44	-0.65	-0.15	-0.24	-0.19
$k + c - z - p = q - a + m$	-0.29	-0.35	-0.33	-0.43	-0.6	-0.89	-0.58
$6731^{\frac{z}{j}} = \frac{87x + o}{zp}$	0.22	-0.35	-0.07	-0.22	-0.46	-0.14	-0.19
$52j\frac{567t}{sd} = 3124/149j * p$	-0.33	0.21	-0.05	0.18	0.36	0.62	0.19
$14^p - 168/t = 243^r$	0.04	0.04	0.04	-0.42	-0.71	-0.46	-0.71
$-a + q + t - p = -l - d - k$	0.04	0.18	0.11	-0.07	-0.36	-0.78	-0.71
$d - 31 = (57w + 88y)bz$	0.11	0.21	0.16	-0.29	0.09	-0.24	-0.19
$u + k - e - q = l + c + y - g$	0.51	0.14	0.33	-0.07	-0.3	-0.78	-0.45
$\frac{87u + 74p + 18m}{36d - 20p - 199w} = or + \frac{n}{31}$	0.29	1.03	0.67	1.51	1.55	1.8	1.61
$(714t - 97q)/(13j + 49x) = 26l/481$	0.4	1.03	0.73	1.19	1.49	1.15	0.84
$\frac{(24a + 3q)(4z + 343t)}{23c(hr + 45)}$	0.69	0.82	0.76	0.9	1.33	1.26	1.35
$670 - lq = \frac{kp + 14r - 119h}{2mf - 32c}$	0.88	0.74	0.82	0.79	1.07	1.26	1.22
$\frac{47 - 5/m}{128^{z/b}} = \frac{n}{18} + \frac{l}{s} - \frac{14}{vb}$	1.13	1.03	1.09	1.68	0.82	1.04	1.09
$(12j - \frac{45}{v})^{x+32/t} = \frac{n - 234}{xas}$	1.42	1.38	1.42	1.5	0.92	0.83	0.84
$119p + 14t + \frac{343d}{12b} = um/(q + 14)$	1.53	1.42	1.49	1.21	1.21	1.04	0.84
$\frac{3 + \frac{31}{p}}{24xg} = (3b - 179u)^{\frac{q}{t}}$	1.53	1.49	1.53	1.23	0.87	0.72	0.84
$\frac{(3a - 7z + aj)(\frac{v}{b} + sb - nj)}{\frac{34p}{27} - az + 14f}$	1.6	1.49	1.56	2.17	2.39	2.33	2.9
$\frac{htq}{(49z - 12t/g)(256j + 119iz)}$	1.9	1.77	1.85	0.65	1.07	1.15	1.09

Table 11. Average Rank of Equations Based on the Various Metrics.

Equation	C1	C2	Avg	*MJX*	*MJXe*	*SRE*	*AsTeR*
$21m/343g = q$	1	1	1	1	4	6.5	5
$4^a + 29^q = 20$	3	3	2	7	3	5	5
$823^{x+g} = z$	2	4	3	2	1	2.5	1
$p + z - x - a = w$	4	2	4	3	2	1	2.5
$vl = \frac{23g}{124bt}$	5	5	5	4	9	12	12.5
$619t/3124za = 847p$	6	10	6	8	15	18.5	12.5
$-e - x + j = c - z$	7.5	7	7.5	5.5	5	4	5
$\frac{246j * 17h}{dlr} = t$	7.5	7	7.5	9	12	15	16
$(6h - f)/v = 14 + d$	9	9	9	12	13	9	12.5
$14^p/a - 119 = z/p$	15	7	10	11	8	11	8.5
$-o + j + y - r = f$	11.5	11.5	12	5.5	6	2.5	2.5
$1123^{l-z} = \frac{123}{t + j}$	11.5	11.5	12	13	7	13.5	8.5
$496u/sd = 27g * bc$	10	13	12	10	18	16.5	18
$k + c - z - p = q - a + m$	14	14.5	14	14	11	6.5	12.5
$6731^{\frac{z}{j}} = \frac{87x + o}{zp}$	19	14.5	15	17	14	18.5	18
$52j\frac{567t}{sd} = 3124/149j * p$	13	19.5	16	20	20	20	20
$14^p - 168/t = 243^r$	16.5	16	17	15	10	13.5	8.5
$-a + q + t - p = -l - d - k$	16.5	18	18	18.5	16	9	8.5
$d - 31 = (57w + 88y)bz$	18	19.5	19	16	19	16.5	18
$u + k - e - q = l + c + y - g$	22	17	20	18.5	17	9	15
$\frac{87u + 74p + 18m}{36d - 20p - 199w} = or + \frac{n}{31}$	20	24	21	28	29	29	29
$(714t - 97q)/(13j + 49x) = 26l/481$	21	24	22	24	28	25.5	22.5
$\frac{(24a + 3q)(4z + 343t)}{23c(hr + 45)}$	23	22	23	23	27	27.5	28
$670 - lq = \frac{kp + 14r - 119h}{2mf - 32c}$	24	21	24	22	24.5	27.5	27
$\frac{47 - 5/m}{128^{z/b}} = \frac{n}{18} + \frac{l}{s} - \frac{14}{vb}$	25	24	25	29	21	23.5	25.5
$(12j - \frac{45}{v})^{x+32/t} = \frac{n - 234}{xas}$	26	26	26	27	23	22	22.5
$119p + 14t + \frac{343d}{12b} = um/(q + 14)$	27.5	27	27	25	26	23.5	22.5
$\frac{3 + \frac{31}{p}}{24xg} = (3b - 179u)^{\frac{q}{t}}$	27.5	28.5	28	26	22	21	22.5
$\frac{(3a - 7z + aj)(\frac{v}{b} + sb - nj)}{\frac{34p}{27} - az + 14f}$	29	28.5	29	30	30	30	30
$\frac{htq}{(49z - 12t/g)(256j + 119iz)}$	30	30	30	21	24.5	25.5	25.5

Table 12. K-mean Cluster IDs of Equations Based on the Various Metrics.

Equation	**C1**	**C2**	**Avg**	***MJX***	***MJXe***	***SRE***	***AsTeR***
$21m/343g = q$	0	0	0	0	0	1	0
$4^a + 29^q = 20$	0	0	0	0	0	0	0
$823^{x+g} = z$	0	0	0	0	0	0	0
$p + z - x - a = w$	0	0	0	0	0	0	0
$vl = \frac{23g}{124bt}$	0	1	0	0	1	1	1
$619t/3124za = 847p$	0	1	1	1	1	2	1
$-e - x + j = c - z$	1	1	1	0	0	0	0
$\frac{246j * 17h}{dlr} = t$	1	1	1	1	1	2	2
$(6h - f)/v = 14 + d$	1	1	1	1	1	1	1
$14^p/a - 119 = z/p$	1	1	1	1	1	1	1
$-o + j + y - r = f$	1	1	1	0	0	0	0
$1123^{l-z} = \frac{123}{t + j}$	1	1	1	1	0	2	1
$496u/sd = 27g * bc$	1	1	1	1	2	2	2
$k + c - z - p = q - a + m$	1	1	1	1	1	1	1
$6731^{\frac{z}{j}} = \frac{87x + o}{zp}$	2	1	2	2	1	2	2
$52j\frac{567t}{sd} = 3124/149j * p$	1	2	2	2	2	3	2
$14^p - 168/t = 243^r$	2	2	2	1	1	2	1
$-a + q + t - p = -l - d - k$	2	2	2	2	1	1	1
$d - 31 = (57w + 88y)bz$	2	2	2	2	2	2	2
$u + k - e - q = l + c + y - g$	2	2	2	2	1	1	1
$\frac{87u + 74p + 18m}{36d - 20p - 199w} = or + \frac{n}{31}$	2	3	3	4	3	4	3
$(714t - 97q)/(13j + 49x) = 26l/481$	2	3	3	3	3	3	3
$\frac{(24a + 3q)(4z + 343t)}{23c(hr + 45)}$	3	3	3	3	3	3	3
$670 - lq = \frac{kp + 14r - 119h}{2mf - 32c}$	3	3	3	3	3	3	3
$\frac{47 - 5/m}{128^{z/b}} = \frac{n}{18} + \frac{l}{s} - \frac{14}{vb}$	3	3	3	4	3	3	3
$(12j - \frac{45}{v})^{x+32/t} = \frac{n - 234}{xas}$	4	4	4	4	3	3	3
$119p + 14t + \frac{343d}{12b} = um/(q + 14)$	4	4	4	3	3	3	3
$\frac{3 + \frac{31}{p}}{24xg} = (3b - 179u)^{\frac{q}{t}}$	4	4	4	3	3	3	3
$\frac{(3a - 7z + aj)(\frac{v}{b} + sb - nj)}{\frac{34p}{27} - az + 14f}$	4	4	4	4	4	4	4
$\frac{htq}{(49z - 12t/g)(256j + 119iz)}$	4	4	4	3	3	3	3

Table 13. Correlation among normalized complexity values of complexity metrics.

	C1	C2	Avg	MJX	MJXe	SRE	AsTeR
C1	1.000	0.957	0.989	0.882	0.827	0.775	0.805
C2	0.957	1.000	0.989	0.932	0.911	0.865	0.864
Avg	0.989	0.989	1.000	0.918	0.879	0.830	0.844
MJX	0.882	0.932	0.918	1.000	0.948	0.922	0.936
MJXe	0.827	0.911	0.879	0.948	1.000	0.971	0.975
SRE	0.775	0.865	0.830	0.922	0.971	1.000	0.977
AsTeR	0.805	0.864	0.844	0.936	0.975	0.977	1.000

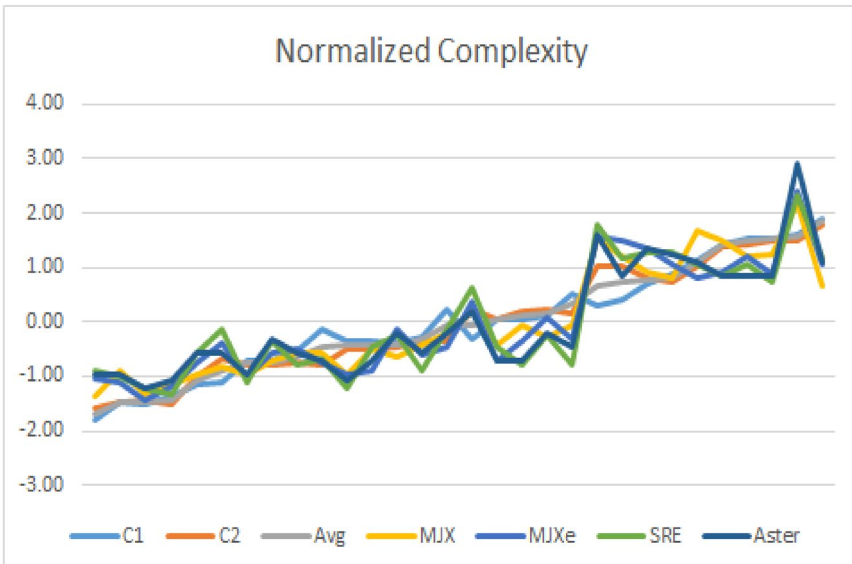

Fig. 5. Variation in Normalized Complexity Values of Complexity Metrics.

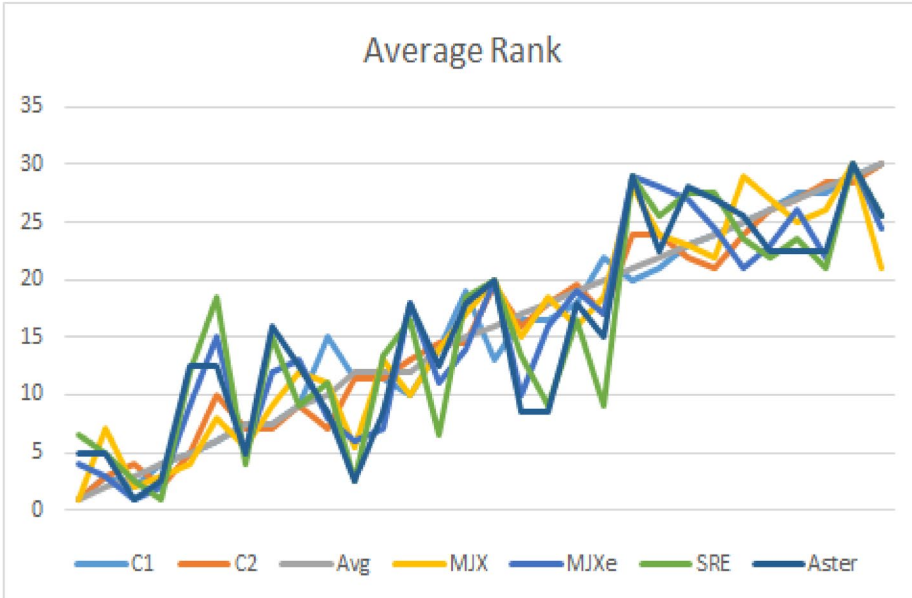

Fig. 6. Variation in Average Ranks of Complexity Metrics.

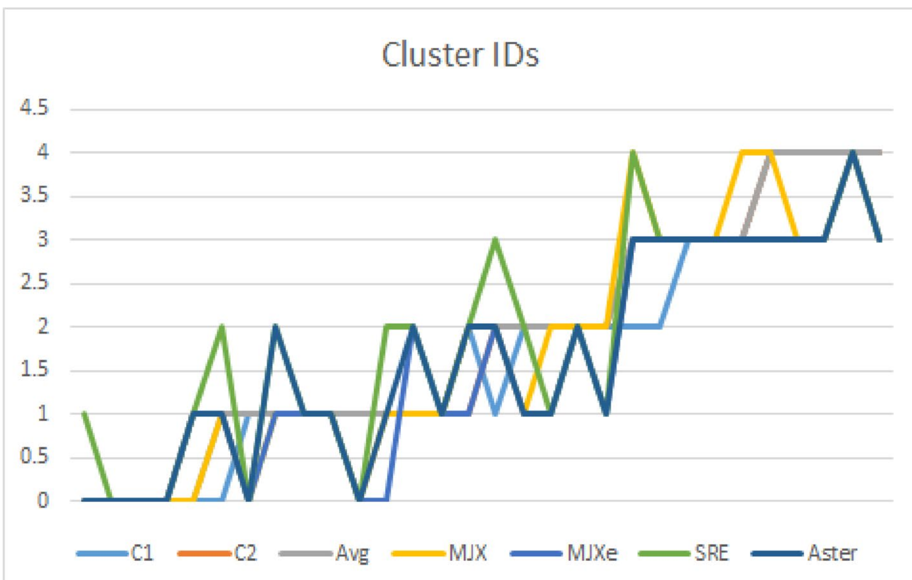

Fig. 7. Variation in Cluster IDs of Equations for Complexity Metrics.

Finally, we computed the correlation among normalized values of the various complexities, shown in Table 13. Here, we observe that all the complexity metrics have shown a high correlation with each other. However, the {***MJX***} complexity metric has the highest correlation (*0.989*) with the Avg. Not only this, ***MJX*** also has the highest correlation with C1 and C2. Finally, we also compared metrics based on the similarity in cluster IDs of the various equations. For this, we calculated the difference between the cluster IDs of an equation across the various metrics. Table 14 shows the count of equations having the same cluster ID difference for a pair of metrics. Here again, ***MJX*** and Avg have maximum (*22*) equations, which have the same cluster ID. Also, ***MJX*** has the maximum equation count, which has the same cluster ID with the C1 and C2. Consequently, the study establishes that MathJax's simple complexity metric is appropriate at least for algebraic equations in an auditory user interface.

Table 14. Frequency Spectrum of Cluster ID Differences for Various Complexity Metrics.

Diff	C1-MJX	C1-AsTeR	C1-MJXe	C1-SRE	C2-MJX	C2-AsTeR	C2-MJXe	C2-SRE	Avg-MJX	Avg-AsTeR	Avg-MJXe	Avg-SRE
-4	0	0	0	0	0	0	0	0	0	0	0	0
-3	0	0	0	0	0	0	0	0	0	0	0	0
-2	1	0	0	3	0	0	0	0	0	0	0	0
-1	4	7	6	6	3	3	1	8	2	3	2	8
0	19	14	13	13	20	18	19	14	22	18	17	14
1	6	9	11	8	7	9	10	8	6	9	11	8
2	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0

Conclusion and Future Work

While most available software for making mathematical formulas accessible offer multiple reading rules and preference settings, they generally leave it to the user to switch as per their requirement. The aim of our work is to build an adaptive system, which can automatically adjust presentation based on a user's level of expertise and the complexity of the equation in question. This work is the first important step in this direction, as it helps us in identifying the best-correlated metric for algebraic equations. Once the complexity of the equation is computed with some confidence it can be easily decomposed into simpler equations while rendering.

Furthermore, this process can also adapt to the user proficiency. All of this is aimed at reducing the cognitive load when a BVI student navigates equations.

Our results demonstrate that the MJX complexity metric has the highest correlation with the various candidate metrics derived from the user study data. It not only shows the highest correlation (*0.932*) with candidate metrics based on the normalized complexity values but also with respect to cluster ID difference. We therefore believe that MathJax's current approach is a valid metric for chunking mathematical expressions into cognitively graspable parts. This is somewhat surprising, given that the metric aims at syntactic and visual complexity while the other metrics take more semantic considerations into account. However, the set of equations we worked with, were relatively simple algebraic equations. Further studies need to validate if this results also hold up for more advanced equations from Calculus, Matrix Algebra, Trigonometry, etc. Also, it needs to be examined that how user background and education influences the results and in particular the cognitive complexity metric. Further, we are aiming to focus on the validation of the interface (hierarchical navigation/Variable Substitution/Other) which can improve the understanding of the complex equations with a significant decrease in the cognitive load. Overall, we see this as a first step in defining a more precise metric and aim to implement it in one of the open source systems mentioned.

The scope of this work is not just limited to the handling of complex equations in auditory user interface. This work will set the foundation for defining the cognitive complexity for a user. This work can also be very well utilized for enhancing the readability of equations in Braille, or verbal description of a diagram. Braille is also a linear representation and hence, comprehending complex equations through Braille is also challenging. Similarly, verbal description of a diagram needs to be adapted based on the context and user characteristics (such as educational background, etc.) and thus this work can also help in contributing to research in that direction.

Works Cited

Bates, Enda, and Dónal Fitzpatrick. "Spoken mathematics using prosody, earcons and pearcons." International Conference on Computers for Handicapped Persons. Springer, Berlin, Heidelberg, 2010.

Cervone, Davide, Peter Krautzberger, and Volker Sorge. "New accessibility features in MathJax." (2016).

Cervone, Davide, Peter Krautzberger, and Volker Sorge. "Towards universal rendering in MathJax." Proceedings of the 13th web for all conference. 2016.

Chang, Lawrence A. "Handbook for Spoken Mathematics:(Larry's Speakeasy)." (1983).

Frankel, Lois, Beth Brownstein, and Neil Soiffer. "Navigable, customizable TTS for Algebra." (2014).

Frankel, Lois, et al. "Development and initial evaluation of the clearspeak style for automated speaking of algebra." ETS Research Report Series 2016.2 (2016): 1-43.

Gellenbeck, Ed, and Andreas Stefik. "Evaluating prosodic cues as a means to disambiguate algebraic expressions: an empirical study." Proceedings of the 11th international ACM SIGACCESS conference on Computers and accessibility. 2009.

Gillan, Douglas J., et al. "Cognitive analysis of equation reading: Application to the development of the math genie." International Conference on Computers for Handicapped Persons. Springer, Berlin, Heidelberg, 2004.

Kacorri, Hernisa, Paraskevi Riga, and Georgios Kouroupetroglou. "EAR-math: evaluation of audio rendered mathematics." International Conference on Universal Access in Human-Computer Interaction. Springer, Cham, 2014.

Maple, Maplesoft. “a division of Waterloo Maple Inc.” Waterloo, Ontario (2016). Wolfram Mathematica, Wolfram Research Inc., Champaign, Illinois, 2009.

MathJax a11y complexity, https://github.com/mathjax/MathJaxsrc/tree/master/ts/a11y/complexity, 2020.

MathJax Consortium. 2019. MathJax Version 3.0. http://www.mathjax.org, 2020.

Murphy, Emma, Enda Bates, and Dónal Fitzpatrick. “Designing auditory cues to enhance spoken mathematics for visually impaired users.” Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility. 2010.

NVAccess, NVDA, Screen Reader, http://nvaccess.org, 2020.

Potluri, Venkatesh, et al. “Significance of paralinguistic cues in the synthesis of mathematical equations.” Proceedings of the 11th International Conference on Natural Language Processing. 2014.

Raman, T. V. “AsTeR: Audio system for technical readings.” Information Technology and Disabilities 1.4 (1994).

Soiffer, Neil. “MathPlayer v2. 1: web-based math accessibility.” Proceedings of the 9th International ACM SIGACCESS Conference on Computers and Accessibility. 2007.

Sorge, Volker. 2019. Speech Rule Engine Version 3. https://github.com/zorkow/speech-rule-engine

Sorge, Volker, et al. “Towards making mathematics a first-class citizen in general screen readers.” Proceedings of the 11th Web for All Conference. 2014.

Stevens, Robert D., Alistair DN Edwards, and Philip A. Harling. “Access to mathematics for visually disabled students through multimodal interaction.” Human–computer interaction 12.1-2 (1997): 47-92.

Exploring the Experiences of UX Professionals in Accessibility

Nandita Gupta, Carrie Bruce
Georgia Institute of Technology
nandita.g13@gmail.com, carrie.bruce@gatech.edu

Abstract

The session will share results from an exploratory research study that delved deeper into the experiences of technology professionals who are interested in digital accessibility. Despite the growing resources within digital accessibility, UX professionals lack support and resources for their accessibility work. This paper builds on these known issues to gain a deeper understanding of the resource needs and experiences of UX professionals as they develop their knowledge and skills in digital accessibility. Our specific aims were to learn more about 1) accessibility practices of digital product teams; 2) motivations, incentives, and rewards for integrating accessibility practices; and 3) resources and supports to develop accessibility knowledge and skills.

Our analysis showed that the study participants reported they had adequate access to implementation and education-based resources but have limited people-based resources. This is important because some participants suggested that having allies and team members who shared their interest or passion for accessibility was a significant influence and crucial support for encouraging accessibility practices. They found interactions with these people-based resources highly valuable and multiple participants indicated a desire for more opportunities to engage with and learn from others.

Keywords

Accessibility, User Experience, UX, people-based resource, support, professional development

Introduction

User experience (UX) professionals including designers, product managers, researchers, and software developers, play a crucial role in accessibility and inclusive design of digital products. They are part of a team responsible for ensuring that a user has the best possible experience with their product, so they must be able to design and deliver usable and useful features that address diverse needs, skills, and abilities. Despite this necessary responsibility, UX and related professionals report multiple difficulties in applying accessibility in their work including limited managerial support (Lazar et al, 2004) and insufficient knowledge (Trewin et al, 2010).

Limited managerial support and lack of accessibility leadership is common as UX professionals are typically employed by technology companies that rarely have a dedicated accessibility department or group of experts. In-house accessibility teams are more likely to be found at assistive technology companies or well-known tech giants. One of the first named accessibility positions was at Microsoft in 1992 and three years later, Microsoft expanded their commitment by developing a corporate policy and plan to provide accessibility features in their products (Schroeder et al, 2012). Since that time, multiple large tech companies, including Adobe, Google, Facebook, Verizon Media, and Oracle, have created specific departments or groups that serve as the contact point for company-based initiatives and resources for implementing accessibility and inclusive design practices. In 2014, the first C-suite positions were appointed at Microsoft and IBM (Miller, 2014). Accessibility and inclusive design are now practiced in various ways throughout digital product companies that must navigate complex legal and compliance issues, user experience requirements, software and testing guidelines, and multiple other aspects of managing business in the technology sector.

While general resources such as conferences, guidelines and certifications are valuable, there is a need for more people specific resources that showcase case studies, stories and in-depth information regarding the field that would demystify the accessibility profession of practice. Professionals within accessibility connect with each other at various events such as meetups, conferences, and networking socials however, these connections are limited to these events which are often paid, and in some case, poorly advertised. Newer UX Professionals who are interested within accessibility are often unaware of resources, as well as various career avenues available within this field. Social platforms such as LinkedIn and Facebook offer various features that enable these professionals to make meaningful connections and learn from one another. However, these platforms have their own set of restrictions. While LinkedIn is a great platform to showcase work, profile as well as build a virtual resume, it relies solely on the use of posts and messages to build connections; there are limited ways to consume information. It does not lend itself towards storytelling and sharing stories and case studies about people unless people share them through their own blogs and posts.

Therefore, while there are ways to teach people through these trainings and programs, along with companies creating support for accessibility practices, the social piece is missing; there is no centralized place to find people and stories within accessibility that are personalized to the role, that would help people make that connection of day to day responsibilities to accessibility. Thus, there is a need to build a broader community of support for accessibility professionals, outside of these conferences and meetups and in addition to these social networks. This study explores the need for these resources, superficially targeting those social resources within accessibility. Creating people-based resources for accessibility professionals would now only create a knowledge base, but also provide a platform for people to seek support and allies.

Methods

To develop a deeper understanding of the accessibility resource needs and experiences, we interviewed and surveyed UX professionals. We targeted their experiences with a range of accessibility resources and practices at their workplaces to better understand the barriers they face in developing skills and knowledge in accessibility. A survey followed after the interviews were analyzed to broaden the pool of respondents and validate the findings from the interviews. Both the interview and survey activities were approved through the Georgia Tech Institutional Review Board prior to the start of research.

Interview Participants

Sixteen professionals working in a wide range of roles related to UX, such as designers, researchers, product managers, engineers, or similar roles participated in this study. We did not ask participants to identify a label for gender. Participants were recruited through accessibility-related conferences (e.g., Disability: IN, Ability Summit, A Future Date), social media (i.e., LinkedIn, Facebook), online groups (e.g., a11y, Systers), word-of-mouth, and personal referrals. To be eligible for the interviews, participants had to be fluent in English, be age 18 years or older, located in the US at the time of the interview, interested in working on accessibility or pursuing a career in digital accessibility, but not currently in a named accessibility position. Participants varied in their level of self-reported expertise in accessibility and years of experience working in technology.

Interview Procedures

The interview covered details about accessibility practices at a participant's current company, their personal motivations for being involved in accessibility work, company-based incentives or rewards for practicing accessibility, and resources or supports available to develop accessibility knowledge and skills. The interview questions were developed based on a structured process of mapping research questions to data needs. We piloted and refined the interview protocol, and developed accessible study materials including consent forms, interview preparation information, and supports for a particular interview (i.e., visual and text descriptions for the Four-Drive Theory). The interview was conducted virtually for all interviews via Microsoft Teams, due to the COVID-19 pandemic that limited in-person interactions.

The interview included a mixture of multiple choice and open-ended questions. We specifically targeted their experiences with three types of resources: 1) implementation-based resources (e.g., style guides, WCAG guidelines, templates, design systems, etc.); 2) people–based resources (e.g., supervisors, consultants, teammates); and 3) education-based resources (e.g., books, webinars, conferences, etc.). Additionally, we asked questions about their accessibility experiences and expertise, and motivations for working on accessibility.

Survey Participants and Procedures

The survey was completed by 21 UX professionals meeting the same eligibility requirements as the interviews. The survey was conducted using an accessible online platform, Qualtrics, and was disseminated through various social media groups and word-of-mouth. The survey questions were similar to the interviews but were mostly formatted as multiple-choice options to validate the findings from the open-ended interview responses.

Analysis

We collected both quantitative and qualitative data from the interviews and surveys. The data was processed and analyzed using Miro, Microsoft Excel, and Qualtrics data analysis tools. We performed a content analysis on qualitative data to uncover relevant findings based on our research questions. Interview notes were converted into virtual sticky notes and arranged in an affinity map to reveal emerging patterns in the data.

Each sticky note was coded to represent various characteristics of the referenced participant. For example, color was used to differentiate the various roles: designers were green, product managers were purple, researchers were green, and engineers were yellow. We used the tagging feature to indicate 'years of experience', 'accessibility expertise', 'company size', and 'other roles'. Organizing the data in this visual manner helped us to find similarities and other patterns related to participant characteristics.

Results and Discussion

The results from the interview and survey validated our initial hypothesis for our research study. This section covers the high-level insights that were uncovered during data analysis: 1) Experience, Confidence and expertise in accessibility practices and the relationships between them. 2) Motivation to pursue accessibility practices. 3) Resources used by professionals within accessibility.

Experience, Confidence and Expertise in Accessibility Practices

We explored the relationship between confidence and experience in relation to implementing accessibility. The following Figure 2 shows the relationship between the self-rated values for expertise level and confidence level for implementing accessibility. The chart indicates that these values were proportional in most cases, but it is interesting to note that a couple of users rated themselves much higher on the confidence scale, despite their lower rating in expertise.

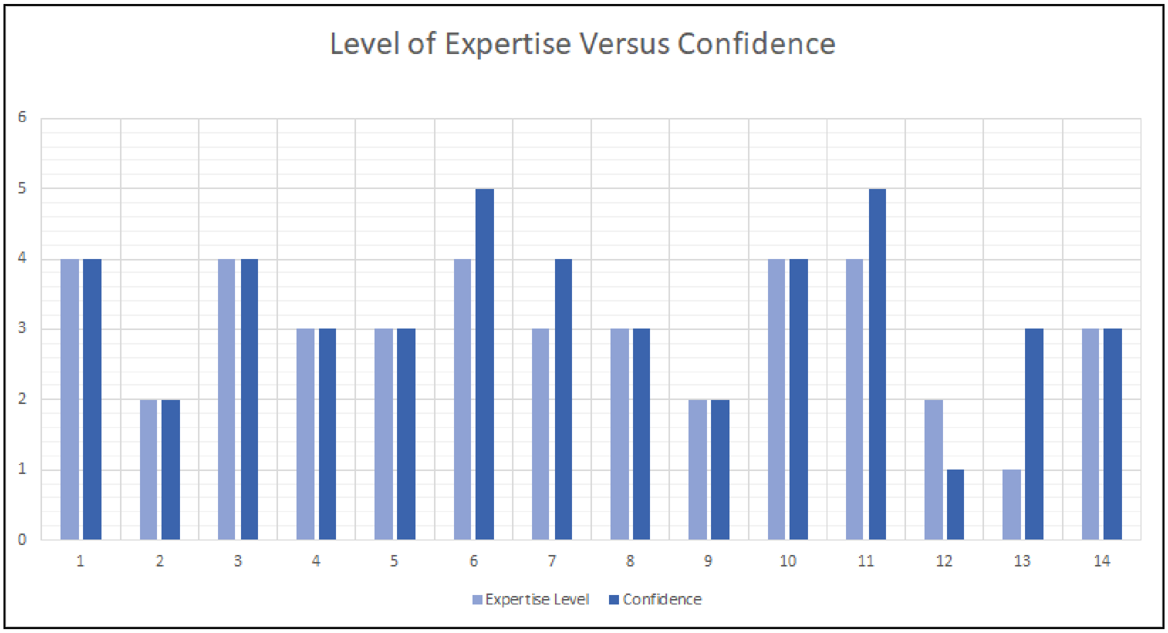

Fig.1 Self-Rating for Expertise Level and Confidence.

Users' confidence on integrating accessibility practices into their current workflow was influenced by prior knowledge and practical and educational experiences. We saw a relationship between the self-rating for confidence level with their prior knowledge within accessibility. Participants who had pursued higher education and certifications, specifically with a focus on accessibility, felt more empowered and rated themselves higher on a confidence scale. In contrast, participants who had limited experience and lacked accessibility specific qualifications rated themselves lower on the confidence scale. One interesting trend within the responses for confidence rating showed that teaching and knowledge sharing experiences boosted the users' confidence; additionally, participants even admitted to the imposter syndrome and attributed their lower rating toward it.

Users' level of expertise on integrating accessibility practices into their current workflow was also influenced by prior knowledge and practical and educational experiences. Analysis showed the various factors that influenced the level of expertise rating for the majority of users: 1) Prior knowledge within accessibility. 2) Work experiences. 3) Training and educational experiences specific to accessibility. 4) Teaching experiences for accessibility. Other factors that influenced this rating were the imposter syndrome, and lack of time commitment towards accessibility. It was interesting to note that users brought up the imposter syndrome again during the self-rating for level of expertise. One user mentioned how their supervisors' perception of expertise imposed upon their personal perception within their self-rating.

Motivation

The users were asked to select their primary and secondary drive for motivation for implementing accessibility practices. These various categories for the drives were taken from a Harvard Business Review case study on employee motivation. The drive to acquire is tied to acquiring goods, things, or experiences. The drive to bond is tied to extending connection towards people, groups, or collectives. The drive to comprehend is tied to experiencing delight in challenges and curiosity for learning. Finally, the drive to defend is tied to the need to defend our ideas, beliefs, or people (Nohria et al, 2008). These brief descriptions were provided to users and they were asked to provide their primary and secondary motivation for pursuing accessible practices. The drive to comprehend and drive to bond scored highest across all interview as well as survey responses; most participants included these within either their primary or secondary motivation for pursuing accessibility practices at their workplace.

Results highlighted that negative experiences and detractors around implementing accessibility practices had different interactions with motivation to pursue accessibility. For some participants, there was an adverse impact on people's motivation towards working on accessibility practices; they were discouraged and felt de-motivated to pursue accessibility initiatives because of these bad experiences. For others, they expressed the need to push harder due to the negative experiences and were more motivated to overcome the opposition. The survey showed the negative effects of detractors on motivation to pursue accessibility practices and users felt demotivated.

Resources

We asked users about the various resources that were provided by their organization to support accessibility practices at work; majority of users found the need for both implementation based and people-based resources, however, during the qualitative interviews, we found more pain points and gaps within people-based resources. Most users felt supported by their organizations for education-based resources.

This data, in addition to the qualitative data analysis from the interviews showed the need for people-based resources. There was a positive relationship between allies and people available to these users and their motivation level for pursuing accessibility. Majority of users who had access to people-based resources more motivated towards working on accessibility practices. They valued various interactions such as activities, sharing resources, lunch and learns, and supporting each other in different ways. They also expressed the need to communicate and engage with these allies in diverse ways and mediums.

Users further talked about the value of these interactions to them; they felt validated, supported and found value through learning with other allies. However, users who lacked allies and mentors within accessibility, expressed a need for interactions with allies and mentors. Majority of users expressed the need to participate in various activities, have discussions about accessibility practices, ask questions and learn from one another and collaborate on accessibility projects and initiatives. They wanted to participate in these activities with allies and their teammates and sought these people-based resources within accessibility.

Additionally, when users were asked about presence of mentors within accessibility, most users felt the need for additional mentorship and support. Users who had allies, expressed the need for mentors within the field. Fig. 3 shows the presence of mentors available to users (from the survey). Most users in the interviews expressed the lack of allies, mentors, and people-based resources. Many users (10 out of 14) mentioned that they faced numerous barriers for support, allies and mentors and their organization around accessibility. Majority of users (9 out of 14) lacked any allies or teammates around accessibility and the few who had allies, mentioned the lack of mentors for accessibility.

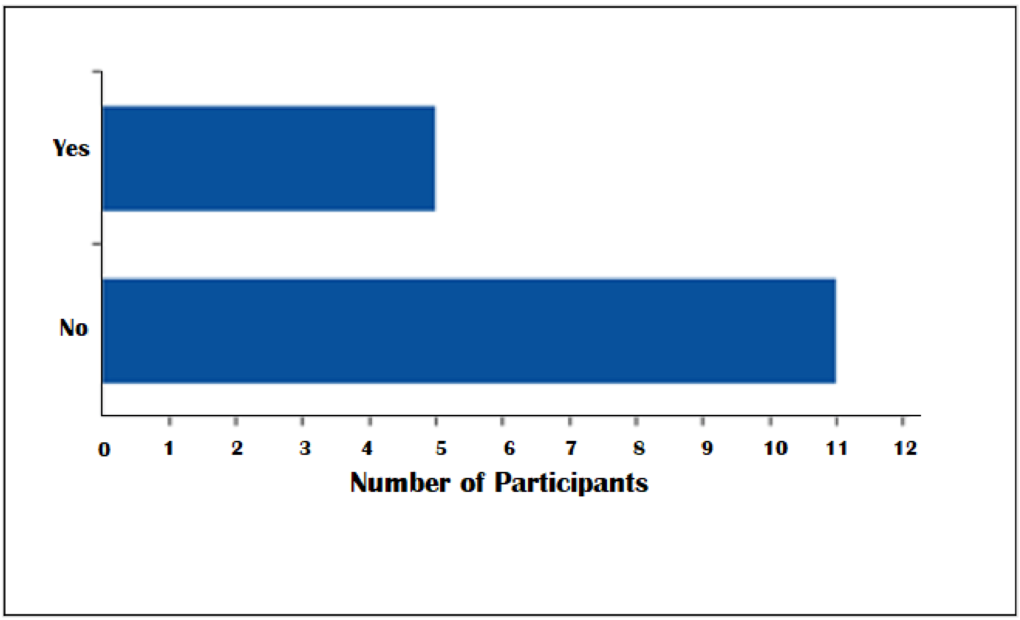

Fig. 2. Chart showing the need for mentors as expressed by users in the survey.

Conclusion

While there are diverse resources available in the industry, there is a need to provide relevant resources and support for UX professionals pursuing accessibility. These resources could be provided in different forms such as interactive guides for accessibility implementation or a social media platform to connect with other accessibility professionals in the field. Currently, platforms like LinkedIn offer limited capacity to connect with like-minded professionals, and other organizations like IAAP require membership to access and connect with other members. These additional support systems and resources would not only motivate other UX professionals to continue working within the accessibility field, but also create a pipeline to introduce other UX professionals to accessibility and inclusive design processes.

Future Work

There are various solutions that may stem from this research; future work for this project will create a people-based resources to support these professionals working within accessibility. There is need to create content in various digestible ways, to engage the accessibility community and build meaningful relationships.

Works Cited

Lazar, Jonathan., Dudley-Sponaugle, A. and Greenidge, K.-D. 2004. Improving web accessibility: a study of webmaster perceptions. Computers in Human Behavior, 20, 269–288.

Miller, Mark (July 22, 2014). "Frances West Announced as IBM Chief Accessibility Officer". *Interactive Accessibility*.

Nohria, Nitin. Groysberg, Boris. Lee, Linda. "Employee Motivation: A Powerful New Model." *Harvard Business Review,* July-August 2008. Case Study. https://hbr.org/2008/07/employee-motivation-a-powerful-new-model

Putnam, Cynthia, Kathryn Wozniak, Mary Jo Zefeldt, Jinghui Cheng, Morgan Caputo, and Carl Duffield. "How do professionals who create computing technologies consider accessibility?" In *Proceedings of the 14th international ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '12)*. 2012, pp: 87–94.

Schroeder, Paul. A Brief History of Microsoft and Accessibility. *AccessWorld*. July 2000. https://www.afb.org/aw/1/4/16165.

Trewin, S., Cragun, B., Swart, C., Brezin, J., and Richards, C., 2010. Accessibility challenges and tool features. Proceedings W4A2010 (Raleigh, NC, USA, April 26-27, 2010).