Designing TADA: Touch-and-Audio-based Diagram Access for People with Visual Impairments

by

Yichun Zhao
B.Sc., University of Victoria, 2021

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

in the Department of Computer Science

We acknowledge and respect the lək̓ʷəŋən peoples on whose traditional territory the university stands and the Songhees, Esquimalt and W̱SÁNEĆ peoples whose historical relationships with the land continue to this day.

Designing TADA: Touch-and-Audio-based Diagram Access for People with Visual Impairments

by

Yichun Zhao
B.Sc., University of Victoria, 2021

Supervisory Committee

Dr. Miguel Nacenta, Supervisor
(Department of Computer Science)

Dr. Alex Thomo, Co-supervisor
(Department of Computer Science)

Dr. Sowmya Somanath, Departmental Member
(Department of Computer Science)

ABSTRACT

Diagrams are graphical representations of data and relations that rely on visual access, which poses challenges for people with visual impairments. In this thesis, we explored how people with visual impairments access diagrams by interviewing 15 participants about their experiences and challenges with diagrams. We found the factors affecting diagram accessibility and conceptualized a ladder of diagram accessibility. We also designed and implemented a novel software system, Touch-and-Audio-based Diagram Access (TADA), based on design goals and principles derived from study findings and existing literature. TADA runs on a multi-touch tablet and enables people to interact with diagrams and have spatial awareness of diagram elements. It supports various interaction techniques using touch gestures and speech as input, and musical tones and speech as output, allowing people to explore, search, navigate, and filter in node-link diagrams. We then conducted a system evaluation study with 25 participants to assess the effectiveness and perceived workload of using the system. The results showed that TADA provides spatial awareness, achieves higher levels of diagram accessibility, and reduces barriers to system access. The thesis contributes to the field of accessibility by providing insights into diagram accessibility factors and needs, introducing TADA with new interaction techniques and audio feedback designs for diagram access, and presenting empirical evidence of the effectiveness and perceived workload of TADA.

Contents

Supervisory Committee ii

Abstract iii

Contents iv

List of Tables vii

List of Figures viii

Acknowledgements ix

Dedication x

1 Introduction 1

2 Related Work 5
- 2.1 Interaction Techniques for Accessibility 5
- 2.2 Making General Graphics Accessible 6
- 2.3 Making Maps Accessible 7
- 2.4 Making Statistical Charts Accessible 9
- 2.5 Making Node-link Diagrams Accessible 10
- 2.6 Summary 12

3 Understanding How People with Visual Impairments Access Diagrams 14
- 3.1 Methods 15
  - 3.1.1 Participants 15
  - 3.1.2 Study Procedure 18
  - 3.1.3 Data Analysis 19
- 3.2 Results 20

3.2.1 Motivations behind Diagram Accessibility and Importance of Diagrams . . . 20
3.2.2 Factors Affecting Diagram Accessibility . . . 20
3.2.3 Unnecessary Diagrams . . . 30
3.2.4 Summary of Answers to Research Questions . . . 30
3.3 Insights . . . 31
3.3.1 A Framework to Incorporate the Findings: Ladder of Diagram Accessibility . . . 31
3.3.2 More on the Relationship Between the DKIW Pyramid and the Ladder of Diagram Accessibility . . . 37

4 Designing a System Enabling Touch-and-Audio-based Diagram Access: TADA 39
4.1 Design Considerations . . . 39
4.1.1 Design Principles . . . 40
4.1.2 High Level Design Choices . . . 41
4.2 System Design . . . 43
4.2.1 Audio Design . . . 44
4.2.2 Supported Interaction Techniques . . . 44

5 Implementing TADA 52
5.1 Diagram Module . . . 52
5.2 Multi-touch Module . . . 53
5.2.1 Fine-tuning the Threshold Values with Informal Pilot Studies . . . 55
5.3 Controller Modules . . . 55
5.3.1 Touch Controller(s) . . . 55
5.3.2 Diagram Controller . . . 57
5.4 Audio Modules . . . 57
5.4.1 ChucK Programs . . . 57
5.4.2 Speech Engine . . . 58

6 Evaluating TADA 59
6.1 Methods . . . 59
6.1.1 Participants . . . 60
6.1.2 Apparatus . . . 63
6.1.3 Study Procedure . . . 65

6.1.4 Data Analysis 69
6.2 Results 70
6.2.1 Task and Question Outcomes 70
6.2.2 Perceived Workloads 73
6.2.3 Participant Feedback 74

7 Discussion 78
7.1 Discussion of Evaluation Results 78
7.1.1 Limitations of the Evaluation Study and Future Studies 79
7.2 Meeting the Design Goals 80
7.2.1 DG1: Achieving Higher Levels on the Ladder of Diagram Accessibility 80
7.2.2 DG2: Reducing Barriers to System Access 82
7.3 Design Approaches and Trade-offs 84
7.3.1 Benefits and Challenges Associated with Preserving Spatial Information 84
7.3.2 Balancing between Information Access and Information Overload 85
7.3.3 Balancing New Concepts with Prior Knowledge 86
7.4 Similarities and Differences with Other Modalities and Formats of Access 87
7.5 Making Diagram Data or Visualization Accessible 88
7.6 Limitations of the Current Design of TADA and Future Improvements . 89
7.6.1 Functionality Improvements 89
7.6.2 Usability Improvements 90
7.6.3 TADA in the Wild 90
7.6.4 Scalability 91
7.6.5 Adaptability 92

8 Conclusion 94

Bibliography 95

A Formative Study Procedure 106

B Evaluation Study Procedure 109

List of Tables

Table 3.1 Details about participants from the formative study. 17

Table 6.1 Details about participants from the system evaluation study. 62

List of Figures

Figure 1.1 Examples of diagrams 4

Figure 4.1 Interaction techniques supported in TADA 51

Figure 5.1 System architecture of TADA 53

Figure 6.1 3D-printed tablet overlays on the tablet used in the study 63
Figure 6.2 Diagram designs to be used as stimuli 64
Figure 6.3 Diagrams presented in TADA as stimuli 66
Figure 6.4 Correctness of answers to quantitative questions listed in Section 6.1.3 71
Figure 6.5 Box plots of NASA-TLX perceived workload ratings. 73

Figure 7.1 Complex diagram examples 93

ACKNOWLEDGEMENTS

I would like to thank:

My mother, Aiping, for always supporting me an ocean away.

My partner, Lalia, and our cat, Jack, for accompanying me through this journey.

My supervisory committee, Miguel, Sowmya, and Alex for the mentorship, support, encouragement, and patience.

UVic, VADA NSERC CREATE, and my supervisors for funding me with scholarships and financial support.

DEDICATION

To the community of people with vision impairments,
and to the research community in the area of accessibility.

May our work contribute to the ongoing efforts
to create a more inclusive and accessible world for all.

Chapter 1

Introduction

Accessing information about the world, such as charts, transportation maps, and educational diagrams, often relies on visual channels. This limited mode of access excludes individuals who are blind, have low vision, or experience temporary vision loss from gaining important and relevant information and can hinder their ability to accomplish everyday tasks at work or in their personal life [65, 49]. To address this issue, researchers and developers have introduced tools and technologies to improve accessibility, including screen-readers [90, 13], tactile graphics [39, 52], and multi-modal interfaces [91, 70, 65]. Our work aligns with these efforts and aims to explore how people with visual impairments can perceive node-link diagrams with improved accessibility.

A node-link diagram is a visualization technique wherein individual data points are represented in the form of a visual point (called a node) and the relationships between the data points are described by connecting the nodes using lines (called links) [95, 82]. For example, it is common to use nodes to represent people in a social network and links to represent their connections or friendship. Other examples include computer networks and character interactions in a book such as those shown in Figure 1.1. People encounter them fairly often in their everyday lives but access to them is often limited [65]. To improve accessibility for such node-link diagrams (and beyond), researchers have conducted studies and proposed new tools.

For example, some researchers have suggested using keyboard input alongside screen readers to access diagrams [22, 13, 46]. However, navigating the diagram per node or connection using arrow keys restricts people to relative spatial awareness, making it difficult to gain an overview and synthesize complex information. Special-purpose tools have also been developed for specific types of diagrams [91], such as chemical molecules [19] and UML diagrams [64], but these were not extended to more general-purpose diagrams limiting their generalizability to other node-link diagrams. Other approaches focus on automatically converting diagrams into text descriptions, but these descriptions suffer from the same limitations as alternative text or image descriptions [30, 45, 50], in that they could require a lot of mental effort, might not be consistent or standardized, and might not capture the full details [50]. Additionally, researchers have proposed using multi-modal access to diagrams [70, 80, 37], but these often involve one-off devices that may not be easily scalable for public use. Lastly, researchers have suggested the use of touchscreen devices [74, 65, 89, 38] to improve access because they are affordable and widespread, and can be used in a mobile situation [31] with audio being the most common output method. To use audio to effectively convey information, interactions need to be designed to direct user attention around a diagram while listening to audio cues [58]. We extend such work using touch and audio by combining dynamic and distinct auditory information and a variety of interaction techniques that utilize touch gestures and speech commands to support multiple levels of information access in our work. We focus on an affordable, widely available commercial tablet, unlike other systems that require specialized feedback hardware. Additionally, we aim to maintain the spatial arrangement of the diagram, giving people a complete understanding of the diagram components and their relationships.

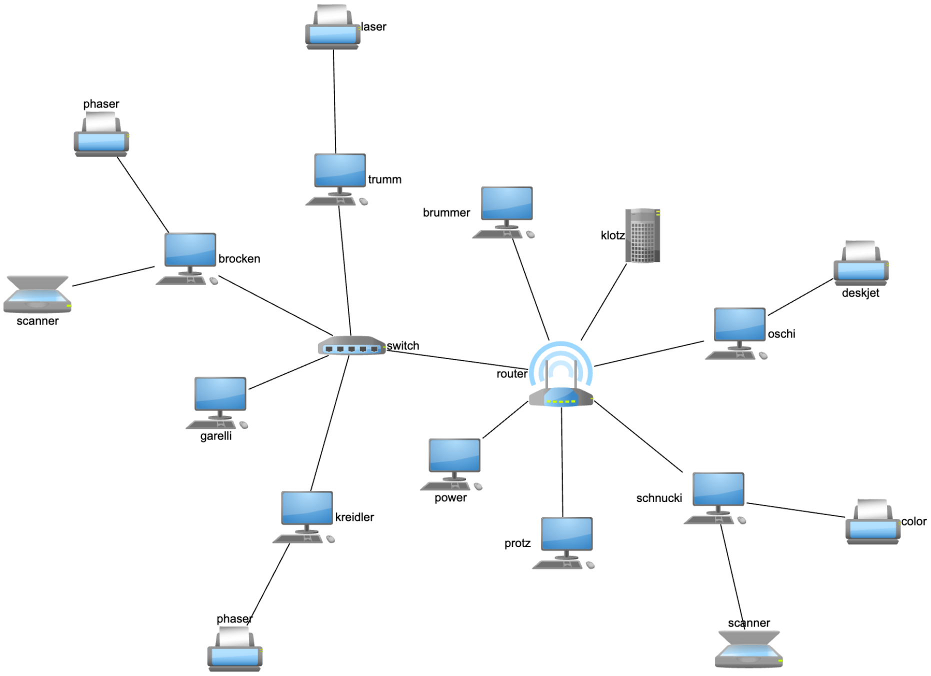
In this paper, we explore the use of readily available technologies and capabilities, such as tablets and audio, to enable access and exploration of node-link diagrams. We conducted a formative study, interviewing 15 participants with visual impairments to understand their experiences accessing diagrams and assess the need for a technology solution to improve accessibility. Based on the insights gained, we identified several design principles and implemented a software system called "TADA" (Touch-and-Audio-based Diagram Access) on a commercial tablet with a multi-touch screen. We scoped the current design of TADA to read diagrams in an existing semantic data format, excluding the process of extracting data from diagrams in the wild [57, 60, 76]. TADA allows node-link diagrams to be accessed through touch and speech as input, with audio as output. By associating nodes and links with distinct sounds and utilizing gesture-based interactions, individuals can accomplish common diagram-related tasks, such as gaining an overview, understanding details, navigating, filtering, and searching for specific node(s) [6, 68]. We then conducted a system evaluation study with 25 participants to assess the effectiveness and perceived workload of using TADA.

In this paper, we make three main contributions:

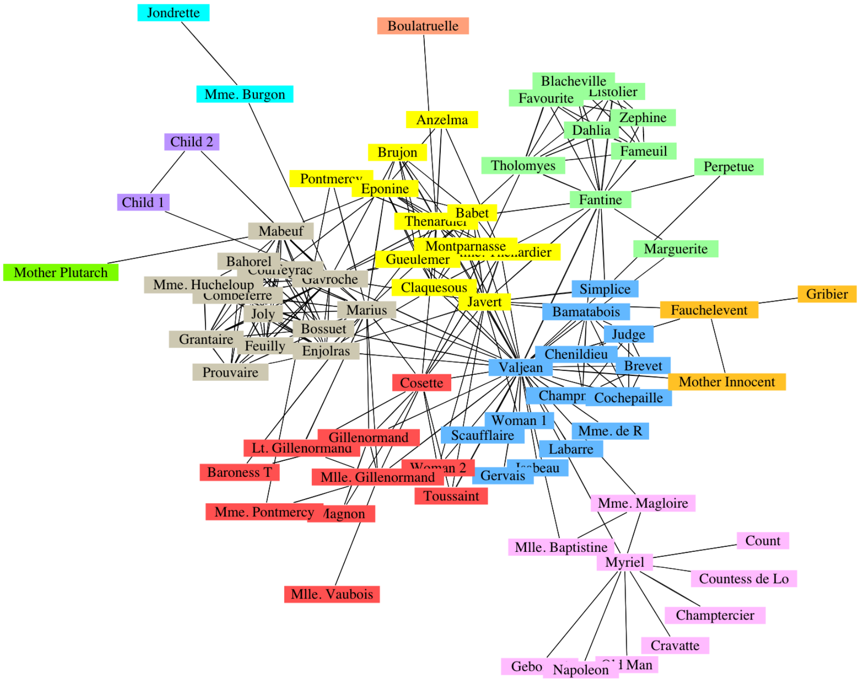
- Insights from our formative qualitative study, highlighting the everyday challenges individuals with visual impairments face when accessing diagrammatic information, limitations of current tools, and the need for more effective solutions to meet different accessibility needs.

- Design of TADA, and a list of design principles that can guide future tool development for improving diagram accessibility.
- Insights from the evaluation of TADA, including its strengths, limitations, and potential strategies for future improvement.

Overall, our work aims to bridge the accessibility gap in accessing node-link diagrams for individuals with visual impairments, empowering them to access and interact with important information at various levels of detail.



(a) Computer network generated using yEd tool [98]



(b) Interactions between characters in Les Misérables [75]

Figure 1.1: Examples of diagrams

Chapter 2

Related Work

We present work related to the central problem of diagram accessibility for people with visual impairments. We discuss literature related to general interaction techniques for accessibility, and then organize other related work according to diagram types and their associated modalities and representations. Due to the complexity of how node-link visualization can be used for multiple types of diagrams, we decide to separate them to be their own categories: general graphics, maps, statistical charts, and finally node-link diagrams. It is important to note that although maps and statistical charts can be a type of node-link diagram, they are usually referred to as separate types of visualization. We lastly summarize the related work and contrast them with our proposed system.

2.1 Interaction Techniques for Accessibility

Generally, guidelines have been proposed for designers to make general touchscreen interfaces more accessible for people with visual impairments. For example, the studies by Tennison et al. [89] explored how people can perceive and follow lines on a multimodal touchscreen tablet using vibrations and sounds. They proposed guidelines for the presentation of non-visual graphics on touchscreens, such as using common tracing strategies, multitouch and anchoring techniques, and horizontal orientations to enhance line perception and exploration. Different feedback signals for start and end points, borders, and inflection points of lines are also important. McGookin et al. [71] derived guidelines for touchscreen accessibility from the evaluation of using two touchscreen MP3 players running on personal digital assistants (PDAs). The study compared one with a raised paper overlay to another one with touchscreen gestures, using sighted participants who were blindfolded. The guidelines include avoiding gestures with a short duration, providing spatial awareness, providing a tactile home key and different button shapes, and providing feedback for all actions. These guidelines can help designers create more effective interaction techniques for visually impaired users.

Kane et al. [53] developed a set of multi-touch gestures for a phone-sized touchscreen device that include a one-finger scan to browse lists, a second-finger tap to select items, a flick gesture to perform additional actions or switch between pages, and an L-select gesture to browse hierarchical information. If music is playing, a double-tap gesture pauses it. The techniques were designed to allow risk-free exploration, reduce demand for selection accuracy, quick browsing and navigation, intuitive gestural mappings, and easy query of location and return home. More recently, AccessWear [59] has a gesture recognition system designed specifically for blind people on a phone, and has a gesture replacement system that allows people to map smartwatch gestures to touchscreen gestures without changing the apps. For example, flicking wrist with a smartwatch maps to swiping right on the phone.

Zhang et al. [99] used an artificial neural network to classify how people who are blind explore digital images with tactile sensations using a haptic device. They identified five exploration procedures: Frame Following, Contour Following, Surface Sweeping, Relative Positioning, and Absolute Positioning. These procedures are characterized by spatio-temporal patterns that are invariant to translation, rotation, and scale. Understanding these behaviours can inform the design of effective interaction techniques using other modalities than tactile.

2.2 Making General Graphics Accessible

Touch screen devices such as phones and tablets can provide audio and haptic feedback which can be utilized to make graphics in general accessible. For instance, GraVVITAS [37] is a tool that uses a touch screen tablet PC to display general graphics (including tables, line graphs and floor-plans) in a visual form for sighted users, while providing haptic feedback for blind users using specially designed hardware and gloves with vibration actuators. The system allows users to explore, inquire, and get an overview of graphics through the use of haptic and audio feedback.

A system developed by Nair et al. [74] focuses on image explorations on smartphones and consists of three tools: a menu and beacon tool, a hints tool, and a quadrant zoom tool. The system features an audio menu that lists all areas within an image, an audio beacon that directs users to their chosen area, an enhanced menu that highlights recommended areas and the number of sub-areas, a prominence indicator that adjusts the volume of spoken area names based on their visual prominence, and a first touch indicator that alerts users when they touch an area for the first time. A11yBoard [100] consists of a web-based drawing canvas mirrored on a mobile touch screen device from a computer to navigate artboards (eg. slide design tools) spatially using touch and gesture while receiving audio feedback and utilizing speech recognition for input and output. The system supports touch-based interactions such as selecting objects, obtaining object information, and starting speech recognition through gestures like split-tap, dwell, flick, and double-tap. Audio feedback is provided to assist users in comprehending the spatial arrangement of the artboard and the system's status.

Another approach is the use of 3D-printed tactile overlays with audio annotations for touchscreen graphics. TacTILE [42], for example, is a toolchain that enables the creation of 3D-printed tactile overlays with audio annotations for graphics such as graphs, images and maps. The interactive app allows blind users to explore the graphic using the tactile overlay and hear audio annotations by tapping the cutouts. Similarily, Maćkowski et al. [70] developed a system that uses a tablet and mobile app to provide audio descriptions of elements on a tactile picture overlaid on the tablet.

Lastly, Swaminathan et al. [86] designed Linespace based on a large tactile display which uses 3D-printing to create tactile graphics. It allows people to explore, create, and manipulate graphical content with touch and speech.

2.3 Making Maps Accessible

Ducasse et al. [31] presented a comprehensive survey of accessible interactive map prototypes for visually impaired users, divided into Digital Interactive Maps (DIMs) and Hybrid Interactive Maps (HIMs). DIMs refer to purely digital representations of maps, and HIMs combine both the digital and physical.

DIMs: DIMs can utilize touch screen devices with vibration and speech feedback to present geospatial information. For example, the TouchOver map [79] uses a smartphone with vibration and speech feedback to present geospatial information from OpenStreetMap data. The user can explore the map with one's finger, and when the finger touches a road, the system vibrates and reads out the name of the road continuously. Sonification interfaces can also be used to convey geometry and spatial information on touch-screen mobile devices. For example, Timbremap [85] is a sonification interface for visually impaired users to explore maps and floor plans on touch-screen mobile devices. The system uses two sonification modes, line hinting and area hinting, to convey geometry and spatial information.

On a tablet with haptic support, GeoTablet [84] is an application designed for visually impaired individuals that offers vocal, auditory, and haptic feedback. The system allows users to explore maps and obtain information about locations and environmental sounds using single-finger contact. Carroll et al. and Lazar et al. designed a weather map application [23, 67] for blind users that incorporates touch and haptic feedback. Users can explore the weather data and zoom more into the map using a touch screen as the input device. The system provides touch-based navigation, auditory feedback through pitch and spatial sound, haptic feedback for points of interest, and text-to-speech for state names. Open Touch/Sound Maps [51] supports haptic, audio, and speech feedback to inform relative information about current position and its surroundings. It uses 3D audio help users estimate the distance from their current location to the next intersection, as well as the direction of the intersection relative to the user's location. The sound's frequency changes with the distance to the intersection, and a higher frequency means closer distance. When the user reaches the intersection, a musical chord is played. Speech is used to announce the name of the current road.

On a large touchscreen, access overlays [54] were designed to improve the accessibility by providing speech and audio feedback. The system supports several interactions including edge projection, neighborhood browsing, and touch-and-speak.

HIMs - Audio-tactile Approaches: Audio-tactile maps are common to make maps accessible as HIMs. For example, Brock et al. [17] developed an interactive audio-tactile map prototype that allows visually impaired people to explore and learn about geographic maps. The system consists of a raised-line paper map overlay placed over a multi-touch screen, a computer, and loudspeakers. The system supports touch-based and gestural interactions for non-visual exploration of maps, allowing users to touch the raised lines of the map overlay to feel the map elements and double-tap on any map element to hear its name and description. Additionally, augmented reality maps can be used to combine tactile, audio, and visual feedback for spatial learning. For example, Albouys et al. [5] designed an augmented reality map prototype that uses tactile feedback such as raised-line maps, magnet boards, and wax-coated woolen strings, as well as audio feedback such as speech synthesis and beeping sounds. The system supports two ways of interacting with the augmented reality map prototype: pointing and pressing, and sliding and holding.

HIMs - Other Approaches: More modalities like the use of speech interaction, keyboard, and other senses can be useful as well. For example, Tikisi [12] is a multimodal framework that supports various interactions with maps primarily through speech commands triggering various actions such as zooming, scaling, moving, centering, finding, and summarizing with touch and keyboard as other input modalities. It uses speech as the output. iSonic [101] is a tool for exploring georeferenced data. It can be used with input devices such as a standard keyboard and mouse, as well as specialized devices like a touchpad, trackball, and joystick. It provides sounds and speech output. Interaction with iSonic includes gist, navigation, situating, filtering, details-on-demand, selecting and brushing. Additionally, the system allows for customization of the information level and the ability to switch between different data views. MapSense [21] is an interactive audio-tactile map for visually impaired children that uses touch, sound, and other senses like smell and taste to help users explore the map augmented using figurative tangibles. These tangibles can contain food and/or scents.

2.4 Making Statistical Charts Accessible

Statistical charts refer to data visualizations such as bar charts, line graphs, and scatterplots. They generally have specialized components and objects such as axes and emphasize data trends.

Text-based Approaches: Alternative text (alt-text) and descriptions can exist for charts, and some systems [25, 73] utilize deep learning models to automatically conduct chart classification, text extraction, and data extraction for visually impaired users to access data in chart images on the web through screen readers. VisText [87], a benchmark dataset for chart captioning, provides three representations of charts: rasterized images, data tables, and scene graphs, and provides different levels of captions for each chart. This would help create chart captions that are rich in meaning and can enhance the understanding, retention, and ease of use of data visualizations. SeeChart [4] deconstructs charts in SVG format on the web and uses a template-based approach to generate summaries of the charts. It follows a data-to-text generation architecture with four stages: data analysis, ranking insights, document planning, and surface realization. The chart is then reconstructed with summary descriptions and interactive features for selecting, navigating, and browsing through the data points are provided via keyboard controls.

Auditory Approaches: Information can be also conveyed through the use of non-speech audio. Infosonics [44] is designed to sonify data sets with aesthetically pleasing audio with two different audio tracks using sonification and speech synthesis. One track generally describes the data and audio encodings, and the other focuses on describing particular data points in speech. Interactions with the audio tracks include play/pause, skip, rate adjustment, and track selection. Chart Reader [90] is a web-based interface that works with screen readers and arranges chart elements in a hierarchical structure, allowing users to navigate through different regions of the chart, including data insights, axes, data points, and filters. Chart Reader also provides textual descriptions and non-speech audio cues for chart elements, as well as options for comparing and filtering series using screen readers.

Question-answering Approaches: A question-answering (QA) system can be utilized to support people in working with visualizations. A system developed by Kim [62] allows users to formulate their queries using natural utterances without needing to interact with interface elements. The system supports interactions such as retrieving values, filtering, computing derived values, finding extrema, sorting, determining data ranges, characterizing data distribution, finding anomalies, clustering, and correlating.

Audio-haptic Approaches: Lastly, an audio-haptic interface can be employed to convey line graph characteristics to blind users. A system developed by Fan et al. [34] uses finger position and contact inclination cues along with sonification and speech output for data exploration. The system supports gist, navigation, and details-on-demand interactions.

2.5 Making Node-link Diagrams Accessible

This section reviews literature which aim to make node-link diagrams accessible for people with visual impairments, using text-based, auditory, tactile, and hierarchical approaches.

Text-based Approaches: Similar to statistical charts, alt-text and image descriptions are common strategies to make diagrams accessible. There exist guidelines [10, 24, 50] and strategies [61, 32, 47] to meaningfully construct the text descriptions, and various ways of automatic conversion from diagrams into text [30, 45, 76, 43, 25]. However, such text descriptions can be limited in their ability to convey information, as they may require significant cognitive effort, fail to convey the richness of visualizations, be insufficient for complex visualizations, and lack consistency or standardization [50]. This can result in perceptual ambiguity and a loss of information [14, 28, 29].

Auditory Approaches: A systematic review on diagram accessibility [91] showed that the auditory approach was predominant to achieve perception. Interaction is not a primary focus in the current research space, and the keyboard is the main way to input information. One approach is to combine touch and audio feedback. Audiograf [58] uses a touch panel and audio feedback to allow users to read simple diagrams and identify diagram elements touched with primarily one finger. Auditory feedback can also come with other input modalities. For example, the PLUMB system [26, 22] displays a graph on a tablet PC and uses auditory cues to guide users through the vertices and edges of a graph. The system supports navigation via a stylus, keyboard, or mouse input, and also provides a command line interface for editing the graph. Similarly, the GSK system [13] is a graph sketching tool compatible with the JAWS screen reader that allows blind and sighted users to create and access graphs as node-link diagrams using keyboard, mouse, monitor, and screen reader input. The tool provides two different views of the same graph: Connection View and Grid View, and allows users to adjust the amount of detail in the auditory cues and use Edge Filtering to select the type(s) of edges. Ramôa et al. [80] evaluated three types of audio navigation user interfaces designed on an audio-tactile device: sonar-based, axis-based, and voice-based, all of which use sound cues to guide the user's hand to specific elements or positions in complex tactile graphics.

Tactile Approaches: Tactility is another commonly used modality [91]. Yang et al. [97] discovered that the preferred tactile representation of network data for people who were blind or low vision was using a node-link diagram, while using list and matrix views was better for adjacency tasks. This broadly aligned with the findings for visual representations as well. The Kevin CASE tool [15] maps schematic data flow diagrams onto a grid-like chart without preserving the original spatial locations of diagram elements. It allows users to read, edit, and create diagrams through tactile overlays, buttons, and speech feedback.

Hierarchical Approaches: Finally, some tools organize information into a hierarchical structure to make it more accessible. For example, the TeDUB system [46, 64, 78, 77] presents technical diagrams such as UML, circuits, and architectural floor plans as a logical hierarchical tree structure of nodes. Users can navigate and annotate diagrams using keyboard or joystick input, and audio, synthesized speech or Braille output. Similarly, a system by Metatla et al. [72] translates relational diagrams into an audio accessible form and organizes information into a hierarchical structure with two presentation modes. Interactions using keyboard commands include browsing, expanding/collapsing nodes, taking shortcuts, shifting perspectives, and requesting navigation information. Kekulé [19] focuses on chemical molecule diagrams and allows users to navigate the hierarchical structure of a chemical molecule using keyboard commands while receiving speech output about the focused diagram.

2.6 Summary

Our proposed system, Touch-and-Audio-based Diagram Access (TADA), builds upon existing work in the field while introducing several novel features that aim to enhance accessibility to diagrams for people with visual impairments. However, our work also differs from past literature in four unique ways:

First, it is specifically designed for node-link diagrams, and allows people to directly query the diagram elements using audio, touch gestures and speech commands. Similar to existing systems focused on node-link diagrams which use speech synthesis and auditory cues [26, 22, 13, 46, 64, 78, 77, 58], TADA also uses a combination of musical tones and speech synthesis to provide auditory feedback for the diagram elements and interactions. In our system, nodes and links have distinct sounds that convey their attributes and types. Interactions have associated dynamic audio cues that help support user tasks, aligning with the guideline of providing feedback for system actions [71]. This audio design allows discrete audio segments to be programmatically combined and scaled up to convey more information depending on the interactions. Together with supported interaction techniques and multi-touch gestures, TADA considers multiple aspects of interaction and provides access to more perspectives of diagram information than most existing touch-and-audio-based solutions [58, 85].

Second, we aim to support a rich set of interaction techniques that leverage touch gestures and speech as input modalities. People can explore, navigate, filter, and query the diagram using a single finger or multiple fingers sweeping, tapping, dwelling, circling, radiating, and using speech commands. One can also combine multiple interactions to help further access diagram information. Existing systems on node-link diagrams mainly rely on keyboard [13, 26, 22, 46, 64, 78, 77, 19] or stylus [26, 22] input which are alternative modalities but could be more limited. We also recognize the benefits of text descriptions [32] and include alt-text support in TADA.

Third, TADA also preserves the spatial layout of the diagram on the tablet screen, allowing people to have an absolute spatial awareness of the diagram components and their relations. A few existing systems either use a hierarchical structure [46, 64, 78, 77, 19] or a grid-like chart [15] to represent diagrams, which serve as alternative ways of viewing diagrams but lose some spatial information as people explore or distort the original layout of a diagram.

Fourth, TADA operates on a commercial tablet device that is affordable and widely available, unlike many other systems that require specialized hardware using tactile feedback [31, 42, 80]. This could reduce barriers to system access and potentially allows users to interact with diagrams in more contexts and environments.

Chapter 3

Understanding How People with Visual Impairments Access Diagrams

Before designing a system to help people with visual impairments access diagrams, we wanted to understand their needs, challenges, and experiences. We did not want to create a system that would be irrelevant, inaccessible, or ineffective for the intended audience. We also did not want to rely on our own assumptions or biases about how people with visual impairments access diagrammatic information. We also thought it would be valuable to seek information which could inform the design of such a system.

As a result, we conducted semi-structured interviews with 15 people with visual impairments remotely through Zoom to understand their experience, challenges, current solutions, and suggestions for accessing diagrams. We had 4 main research questions (RQs) for the interview study:

RQ1: How do people with visual impairments currently access diagrammatic information such as flowcharts, social network diagrams, and organizational charts?

RQ2: To what extent are people with visual impairments facing challenges when they access diagrammatic information?

RQ3: What are the current solutions or workarounds to the challenges that people with visual impairments face when they access diagrammatic information?

RQ4: What are some suggestions or ideas that could benefit future technology design?

3.1 Methods

Here we describe the methodologies for our recruitment, data collection, and data analysis process. We used the format of a semi-structured interview. This would allow us to ask questions more flexibly and spontaneously, and allow the participants to expand upon their specific answers to the questions. We formulated the interview questions from the RQs by breaking them down into more specific and detailed questions for participants.

3.1.1 Participants

We recruited 15 participants who met the following inclusion criteria: they considered themselves to be blind, visually impaired, or partially sighted; relied on the auditory channel or a combination of auditory and other sensory channels to access digital information; used screen readers, braille-based systems, self-created solutions or other methods to access digital information; and were 18 years of age or older. We excluded people who were unable to utilize sound sources for accessing information.

We recruited participants by using three main sources of information: email contacts and mailing lists of various organizations related to visual accessibility and assistive technology, social media advertising on platforms such as Facebook, Discord, and Reddit, and snowball sampling where existing participants recruit future ones through their network. We obtained permission from the potential organizations to publish recruitment materials through them and posted the advertisements on relevant social media channels. Prospective participants who were interested in the study contacted us and were screened for eligibility. We then provided them with the consent form and other study materials and scheduled a time for a remote interview.

The participants' demographic information is summarized in Table 3.1. 8 of them are male and 7 are female. They range from 18 to 69 years old, with an estimated (due to age ranges) average age of 47.7. All of them have encountered diagrams in personal, professional, or educational settings with different frequencies from daily to sporadically. P8, P11, P12, P13, and P14 were able to provide diagram examples to be showcased during their interviews. They were given a remuneration of CAD$30 for each interview which lasted about 55 to 70 minutes.

PID	**Age**	**Gender**	**Degree of Vision Loss**	**Onset of Vision Loss**	**Background**	**Average Frequency of Knowingly Encountering Diagrams**
P1	35-40	Female	Total	Birth	Employee resource group lead	Weekly
P2	38	Male	Total	Young adult	Government employee; Salesperson; Worked in the sports and recreation industry	Quite often, more frequently than weekly
P3	35-40	Female	Total	Birth	Retail	Daily in school(s); Occasionally nowadays
P4	18	Male	Total	Birth	Student	Daily in school(s); Occasionally nowadays
P5	52	Male	Blurry, glasses do not help	Recent years (3 years ago)	Hardware support, IT	Often in course work; Rarely nowadays
P6	32	Female	Total	Birth	Government employee	Daily in school(s); Occasionally nowadays
P7	60	Male	3% vision	3 years old	Teacher	Half a dozen to a dozen times a week
P8	60	Male	Total	7 years old	Businessman	Not often

P9	40	Female	Congenital Rubella Syndrome and Glaucoma. Fluctuating vision	Lifelong progression	Human resources, volunteer resources	Quite often in course work; Not too often in general
P10	63	Male	1% vision	Birth	Banking and assistive technology	Sporadic
P11	69	Male	Total	Early seventies	Self-employed	Not very often (tend to avoid diagrams)
P12	56	Female	Legally blind, low vision	Losing vision for the past decade	Professor	Frequent
P13	62	Female	Total	2 years old	Information officer, now retired	Not very often nowadays
P14	42	Female	Legally blind, significant vision loss	Birth, lifelong progression	Research and teaching	Less than once a week at work; Occasionally in general
P15	50	Male	Total	Vision loss from birth, total vision loss from 2 years ago	Wood products engineer	Not very often (gave up on diagrams generally)

Table 3.1: Details about participants from the formative study.

3.1.2 Study Procedure

Before the remote interview, we asked participants to think about 3 scenarios of encountering diagrams, their experiences with accessing the information, and screen-share the diagram examples during the interview if possible.

At the start of the interview, we inquired about the participants' demographic information, including age, gender, background, and description of vision loss. The rest of the interview was divided into scenarios with respect to the encounters of participants with diagrams.

Within each scenario of a diagram encounter, we asked about the diagrammatic information, context of the encounter, importance of the information, motive(s) of accessing the information, tools participants used if any, and tasks participants performed. We then asked the participants to share things that went well and challenges during their encounters, and workarounds, solutions, and suggestions that they might have come up with to help them better access the diagram.

The detailed study procedure can be found in Appendix A. Here is the list of questions asked for each scenario of a diagram encounter:

1. Describe briefly the diagrammatic information from your encounter.
2. Describe briefly the context of the encounter in terms of when, where, how, who, and why.
3. To which degree were you successful in accessing the diagrammatic information? (from 0, not successful at all, to 5, perfectly successful)
4. How often do you encounter such a scenario?
5. What were the motives for or objectives of accessing the information?
6. What was the importance of the information?
7. Was it an individual, collaborative, or hybrid setting?
8. What was the surrounding physical environment? For example, was it public or private? Was it personal or professional?
9. What media was the information presented in? Was it printed, electronic, verbal, or other?

10. What did you do when you accessed the information? What tools did you use to access the information, if any?
11. What types of tasks do you do with the diagrammatic information? (Some examples include getting a general idea, searching, exploring details, browsing, navigating, and annotating.)
12. What went well and smoothly when trying to access the diagram?
13. What were the main challenges that you encountered when trying to access the diagram?
14. How did you try to address these challenges, if at all?
15. Do you have any ideas that would help you access diagrammatic information in better ways?

3.1.3 Data Analysis

We used thematic analysis with open coding to analyze the data [2]. This methodology was appropriate for our interview and RQs because it provided flexibility and allowed us to be open to finding high-level insights that we might not specifically look for. At the same time, it kept us focused on the topics of interest and relevance, balancing between structure and openness.

We transcribed the interviews and open-coded the transcripts in NVivo [69]. The transcript of P1 was first coded, and the resulting codes were initially grouped. A vastly different transcript was then chosen to be coded next to broaden the possible scope of the codes. After coding these two transcripts, we discussed about refining, adding, and removing codes, and an initial codebook was created and used to code the remaining transcripts. New codes were added as they emerged, and we went back to older transcripts to apply the newer codes. This was an iterative process.

After we finished coding and finalized the codes, we performed affinity diagramming [55] using index cards, post-its, pins, and a bulletin board to further cluster the codes and discover themes that are distinctly separate from the codes and research questions.

3.2 Results

In this section, we explain the results in terms of the importance of diagram information, the factors affecting diagram accessibility, the instances where diagrams may be unnecessary, and how these connect back to the research questions.

3.2.1 Motivations behind Diagram Accessibility and Importance of Diagrams

Participants found diagrams important for several reasons. In work settings, certain diagrams contain crucial work-related information (P1, 2, 7) such as Swim-lane diagrams for project management. In educational settings, diagrams are almost unavoidable, and they help present concepts and supplement other learning materials (P2, 3, 4, 5, 6, 8). In personal settings, diagrams can be accessed for personal curiosity or leisure purposes such as exploring the layout of a geographical area or browsing the internet (P4, 7, 8, 10, 11, 13). Sometimes, participants accessed diagrams because they wanted to verify the information to be relevant and accurate when they wanted to share diagrams with other sighted people (P7, 13, 15). In general, participants found diagrams to be a way to enhance or support existing information and benefit the learning process (P1, 2, 5, 6, 8, 11, 12), present *"specific information that's only [in a] diagram"* (P2) and may not be accessible otherwise (P1, 2, 7, 12), present the spatial locations of objects (P8, 10, 12, 13), and highlight the relationships between objects or systems (P2, 8, 12, 13).

3.2.2 Factors Affecting Diagram Accessibility

One of the main recurrent topics in the participants' answers was the factors that affect whether they could succeed at accessing the diagram information or not. Participants reported a wide range of factors, which we report in three groups: personal differences, the properties of the diagram itself, and the contextual factors of the situation when a diagram was accessed. Personal differences refer to the unique characteristics of the participants, such as their experience with diagrams and tools, different skill sets that benefit themselves in different ways, and different objectives in accessing the diagrams. The properties of the diagram itself refer to how it is produced in terms of accessibility and modality of access, and how the quality and quantity of the accessible information vary which in turn affect participants' perception. The contextual factors of the situation when a diagram was accessed include factors that are not directly related to the participants or diagrams but can affect participants' access to diagrams, such as the availability of accessibility tools, the availability and helpfulness of people who were sighted, and a lack of reinforcement of accessibility standards.

3.2.2.1 Individual Differences of Participants

Participants often described how their personal abilities, experiences and goals directly affected their likelihood of success when accessing information. Sometimes participants also offered insights about how they thought others' abilities and skills would affect them. This diversity in individuals and personal circumstances is a recurrent topic in multiple studies of the blind and low vision community [49, 37, 38, 94], yet it is valuable for us to highlight the particular ways that our participants explicitly identified. We group these in terms of cognitive spatial ability, experience with diagram topics, accessibility tools and modality of access, and goals.

Spatial Cognition: A large number of participants commented on the importance of the spatial information contained in diagrams (N=13). For example, P2 explicitly highlighted that it is crucial to understand the *"items"* in a diagram, the *"relationships of space and the items that take up the space"*, and the *"space in between them"*. This indicates that there is substantial awareness of the importance of the spatial aspects of the representations in diagrams. Yet, participants also commented on how people with different levels of visual perception and, crucially, with different onset of vision loss (from birth, as children or as adults) might encounter significant challenges to access this information. P11 (one of six participants commenting on this) comments that *"I have that spatial awareness. Some blind people do not have [it], especially [those] blind from birth; They don't develop that spatial awareness."* P10 who was born blind further explained: *"It's the hardest thing ... Spatial understanding is a real challenge for blind people because they don't have that learning that children have and young people [have] in looking at maps, understanding scalability, understanding by looking at something, understanding the scale, and then understanding the distance. It's something which often blind people, especially people blind from birth, don't have - there hasn't been a lot of ways for that to be learned."* .

Experience: The level of proficiency with accessibility tools could also influence one's success in accessing diagrams. P15 explained: *"I know there'[re] people out there that have super JAWS [which is a screen reader program] skills and they could probably interpret* *a graph on a web page because there're probably ways of figuring it out ... But a super JAWS user ... that's 2 or 3% of the people that use the program."* Having a higher level of proficiency with tools can give people a higher degree of access, but not everyone has that level.

According to some participants, having prior experience with the background of the diagram could also help with understanding. P2 said: *"I had no problem knowing [diagrams about ergonomics] because of my kinesiology background in school, and all our diagrams are way more complicated, so I had no problems memorizing, or even being able to memorize it so well that I can talk about it."* However, even when someone is an expert in the topic of a diagram, there can still exist a barrier to accessing or understanding it. P1, a professional project manager, struggled to make sense of Gantt charts which are *"painfully inaccessible"*, and Swim-lane diagrams which were *"crucial [but] difficult"*.

Some participants built metaphors of diagrammatic information by associating it with familiar physical objects. P4 commented: *"My first step basically is to try to match whatever diagram I'm feeling to a shape or 3D form that I feel on a regular basis ... If it's a graph, then I would try to picture a flat surface like a wall, and then imagine the x and y axis on that."*

Familiarity with Modalities: Individuals also reported different abilities and preferences regarding the modality used to render information (e.g., audio vs. tactile). P2 (who lost vision later in life) speculated that the use of audio might be better suited for people who are born blind: *"I think people who could see before may have a better chance at doing that, because ... the visual cortex of their brain hasn't been rewired as much as people who were born blind. For people who are born blind, or if there are people who lose their sight earlier, the longer you spend being without sight, the more your visual cortex gets rewired to the audio."* However, P6 (who was born blind) commented that it depends on the individual: *"Sometimes it's just the audio is [kind of] hard ... depends on how each person learns. If it could be more hands-on [and tactile, it] might be helpful to other people."*

Individual Goals: Participants identified that success in accessing diagram information depends on what the individual needs at that particular moment. P12 explained that an inherent limitation of using screen readers is that they could give unnecessary details that the participants do not want or need: *"I don't really need to hear the website every time ... I don't need you to spell out or to list every single researcher ... Can we* *skip that part? Sometimes, you just want to skim something, and you can't. You can't do the "cheaty" short[cut] method of looking at the diagrams [visually] and inferring the information."* With textual descriptions of diagrams, people have to listen to everything in order to look for the pieces of information they are looking for, and this could be time-consuming and frustrating.

P6 used the diagram example of a computer networking diagram and explained that the accessible information might not be enough for tasks which require more details: *"If this had come with an instruction to say, 'I need you to set up the office,' then I probably would have thought: 'oh, I don't have enough information.' ... I think it makes a difference too if I actually need it for something [to perform tasks on], or if it's just kind of something that's just there."*

Participants also highlighted how sometimes the accessible information is not related to what a particular user wants at all. P10 gave a related example: *"In some instances, I could see individuals who are looking at that diagrams who want to understand the design, or they want to understand the layout, and they just don't want the statistical information [which was made accessible] ... they want to understand the lines and the flows and that sort of information. So it's a big challenge."*

The examples above indicate that there can be a mismatch between the way that a diagram is accessible and the particular goals that a participant has. Although some information is accessible, it might not help people achieve their goals effectively or efficiently.

3.2.2.2 Properties of Diagrams

Participants reported that the accessibility of a diagram depends on its complexity, how it is made accessible, and its different representations.

Amount or Complexity of Information in Diagrams: Multiple participants (N=11) explained that when the amount or complexity of the information from a diagram gets too large, they fail to understand the diagram or take a long time and lots of effort to. P2 explained: *"If you put too much information in one thing, it's harder to decipher, or you spend most of your mental efforts targeted towards deciphering. But then, now you miss the understanding."* This is echoed by P4's experience with an alternative text: *"It's so much more wordy or long that while I'm trying to read it, I get lost because I'm trying to absorb the information - important information, and I kind of lose track of that reading, like the huge chuck of the paragraph, and it really doesn't have to be that* *way."* P14 also explained that sometimes visual visualizations add more complexity to the diagrams: *"Sometimes in diagrams, it can be more complicated because researchers might perhaps use colour to signify ... something special ... Some diagrams have that extra layer of complexity with different colours, or the font, or the boxes, etc. and so that ... it makes it just more of a challenge I think."*

However, participants also described some general strategies to prevent the diagrammatic information from getting too overwhelming. Dividing the information into chucks could make the digestion of it easier, as commented by P2: *"[The diagrams] become unwieldy or too much information at once. It's always best to simplify and break up the topic into multiple diagrams, or you risk making it useless."* For tactile diagrams, if multiple attributes need to be shown on a single diagram, P11 explained that having different tactile textures help distinguish different parts of the diagram: *"It's to interpret the lines of the edge of a lake, as opposed to something else, to keep it all kind of segregated or not get the lines confused, and to retain the whole perspective of lakes and rivers and that kind of stuff."*

P8 explained that a related strategy is to gain an overview of a diagram first to understand the general layout, and then gather the details using diagram descriptions: *"Don't zero in on detail first, start with a basic outline and then try filling in the detail."* P15 also confirmed: *"You have to tell the basic first and then you tell the details of it ... You have to draw the picture clearly first, and then you get the information."*

The clarity of the information in diagrams is also influenced by the intersection of the complexity of the diagram and the objective(s) of the audience. As we discussed in Section 3.2.2.1, different participants have different goals in accessing diagrams, which affect how they perceive the relevance and completeness of the information. When a diagram is relevant, keeping the accessible information concise is also important as mentioned by P3: *"I think just keeping it as simple as possible while still conveying the relevant information is really the key to a good diagram."* This is confirmed by P4: *"When I'm trying to listen to a diagram, I'm looking for conciseness ... I don't want it to be an essay. I don't want it to be like long sentences, because I don't care about grammar, or I don't care about how detailed something is. I just care about getting the information in a short and concise manner that I can easily take in without getting overwhelmed."*

Some also managed to create alternative representations of a diagram. P7 said: *"Sometimes you have to come up with new ways of representing things ... so that you know this relates to this, but in an alternative way it relates to that as well."*

How Diagrams are Currently Made Accessible: 14 participants reported encountering diagrams that were either inaccessible or insufficient for their needs. P1 states that: *"In the real world, people don't know how to produce information accessibly, like that is just a battle that unfortunately will never end in my lifetime ... It's just the way it is and people don't know ... what they don't know."* P2 affirmed this experience, explaining: *"Sometimes you might get the alt-text, but the annotation boxes may not be accessible ... or sometimes you get the annotations mixed in with the regular description, but they get pasted in like ... they all ... shuffled in each other."*

Some resorted to using workarounds. P5 said: *"Unfortunately there was no alt-text for the screen reader to [read] ... [because] people didn't construct alternative text to it. So you have to use OCR to try extract[ing] as much information as you can."* P9 showcased how a document containing diagrams was accessed using a screen reader which only managed to tell the user that there is a graphic present *"which means nothing to me"*. P12 also states: *"It tells us everything on the other [textual] pages, and then you hit the page with the diagram, and it stops."*

P2 added that sometimes the quality of the accessible information is poor: *"They are accessible in the way that the computer will read them out automatically as it goes through the web page. But then the understanding isn't always there, because the descriptions are either half-heartedly inputted in or they're not well thought out."* P9 also highlighted the observation that alt-text made for images is better than for diagrams: *"I haven't really found too many graphs in diagrams that were really extremely well done."*

P9 commented on the limitations of new computer vision technology in resolving accessibility barriers: *"AI and stuff like that aren't quite accurate yet."* P7 also confirmed this, stating that *"It doesn't always give you the details ... Mostly, if you're asking what I use them for, it's just so that I know what's in the picture."* While these technologies are being developed to help recognize visuals in a diagram and build a description automatically, the generated descriptions might not be detailed enough or accurate.

In fact, even when a person with vision loss is involved in the design of a diagram, the diagram still gets made inaccessibly. P12 who is a researcher states: *"I'm part of the problem is what I'm saying. I'm part of a group that's creating material and it's not all accessible."*

Nevertheless, sometimes diagrams are well-produced in an accessible way. P4 praised that the university P4 attended provided the needed resources: *"Here in a college setting ... people are a lot more aware of these technologies, so I've encountered a lot more actual descriptions with images and diagrams in terms of getting it in either braille format or a* *described text ... These people are usually expert at this, and they do a very good job at converting these diagrams to an accessible format ... [providing the] details that I'm able to get out of it, and the flexibility - since it's something that's with me, I don't have to keep asking people or anything. [This] is also a plus point."* P7 confirms this experience with alt-text: *"If the human-generated alternative text is done well, then you can gain a very good picture of the diagram."* P3 also shared the experience with well-constructed tactile diagrams: *"It was spaced out really well, and it wasn't just a big clump of lines. It was very easy to differentiate one part from the other."* Hence, when a diagram is made in a way that is not only accessible but is also clear and gives relevant information to the audience, it would contribute to one's success at accessing and interpreting it.

Representations of Diagrams: According to participants, diagrams are usually presented to people with visual impairments in four ways: printed, digital, tactile, and verbal. Each method has its own advantages and disadvantages.

Printed diagrams can sometimes be accessed with visual assistance like a digital magnifier depending on the degree of participants' vision loss. However, as confirmed by P10, it is often difficult, time-consuming, and limited: *"It was just too bulky and too time-consuming, and as my vision went [worse], it was just too difficult because you also get a little motion sick as you're moving the paper around trying to get it lined up underneath the camera ..."* and *"... With paper, the limitation is that there's no ability to get additional information."*

P1 explained the limitations of accessing digital diagrams via a screen reader reading the text description: *"[The alt-text] gave me a sense of what it was about, as alt-text should ... Alt-text is meant to basically give you the 'need to know', but in terms [of what] I wanted to know ... [For example,] where do the different branches [in the diagram] go? And ... what's the relationship? There was no description of that."* This highlighted that digital diagrams, when accessed via a screen reader, only describe what counts as necessary to the designer of the description and do not provide freedom to the participants to choose which pieces of information are to be accessed. P2 commented on the possible technical limitations restricting the amount of detail included in alt-text: *"You can't put in so much description, whether it [is] because of the space, the amount of characters that are allotted to alt-text, or it could be because the description will become too clunky for a person to remember or visualize."*

A description is also a linear way of accessing information and this can be limiting, as confirmed by P4: *"Alt-text is very linear and sometimes diagrams do need to be accessed* *in a way where you need to actually observe the entire diagram itself."*

While some participants reported limitations of alternative text, others also mentioned its advantages. When details might be missing, the description can serve as an overview of the diagram which help with their understanding as mentioned by P14. Tasks such as gathering data points and comparing data could be easier with a description, as commented by P4: *"Collecting data from the diagram ... like values ... is more easily doable with the alt-text"* as compared to using tactile diagrams.

When the descriptions are constructed well, they can be accessed and referenced reliably as many times as P4 wants in P4's coursework: *"I can actually rely on people who've actually kind of made sure that this is an as accurate description as possible, and it's something that I can just rely on and go over as many times as I want, because it's something that's basically at my disposal. So just like people looking at those diagrams, I can just read it or feel it as many times as I want. So that's definitely a plus point."* In a professional setting, however, *"[employers] have to spend more money to make a special text description, and then that oftentimes does not happen because of budget; That's more money"* as explained by P2.

Some participants thought that tactile diagrams could better support tasks such as navigating or observing the entire diagram. P4 said: *"When you actually need to observe the diagram, and you [want to] know what's going on, then I would definitely prefer navigating ... which currently I do to the tactile or braille."* P11 explained: *"[If] it's the tactile access, you can very easily move around, whereas if it's in text which is presented audibly, then you have to work your way through sometimes, which takes longer."* P4 also contrasted tactile diagrams with alt-text: *"Alt-text is very linear and sometimes diagrams do need to be accessed in a way where you need to actually observe the entire diagram itself."*

At the same time, P10 states the drawbacks of tactile diagrams: *"The vast majority of [tactile diagrams] are useless, that they don't provide the reference of information, or that they're too difficult for the users to discern, or there's a lack of standardization in the type of legends that are used for what certain elements are."* P15 commented on the challenges with tactile diagrams often needing to be specially made and costing time: *"[Tactile diagrams] are just not quick enough [to be accessed] ... Somebody had to mail that out to you ... The information should be just as quick to get as it is for somebody that doesn't have a visual impairment."*

When participants could not gather information independently from a diagram, they often result to a verbal description of it from someone else who is sighted. We discuss this more in the next section (Section 3.2.2.3).

3.2.2.3 External Factors from the Larger Environment

Aside from one's abilities and experience, and properties of a diagram, participants also recognized the effect that the environment and context have on accessing diagrams.

Access to Tools or Other Resources Available: Existing tools help people access diagrams, but they might not be accessible due to monetary cost or time cost. P1 explained: *"If there are tools available, they're probably thousands of dollars because anything accessibility-wise is thousands of dollars."*

Some participants mentioned that in an educational setting, there were organizations that specifically help make learning materials accessible by including descriptions or producing tactile diagrams, but similar challenges could apply because the process of making information accessible could be time-consuming. P9 sent a textbook to an organization to convert the images into descriptions but P9 *"[had not gotten] the textbook back yet due to a backlog."* although the aim was to deliver the accessible materials ahead of classes starting.

Access to Sighted People Around, and Their Interpretation: Asking others who are sighted is a common workaround for many when the diagram is not accessible (N=14 of our participants reported this). However, participants also discussed the associated drawbacks of this strategy because other people might not always be available, and it is inconvenient as P1 expressed: *"The more you have to rely on sighted people for something, no matter what it is, it's already inconvenient ..."* and *"... Getting sighted people is a privilege and added premium. I can get them if I need them, but I have to ... weigh my favours."*

Even when sighted helpers were available, this approach could still fail. P1 added that if the other person does not know certain domain-specific knowledge from the diagram, it is still difficult: *"If they don't know the terminology, they're not gonna be able to provide it to you in a way that you're gonna understand."* P11 also commented that the others' descriptions might lack interpretation: *"Even having someone read it for me, they wouldn't be able to interpret it well enough for me to gather any information from it."*

Furthermore, participants highlighted that due to the verbal nature of asking someone else to describe, they had to receive and process the information in real time which can be challenging. P8 said: *"When people try to explain it verbally, it's actually very challenging* *to put it into the spatial position in your brain."* hindering the understanding of a diagram as a result.

Participants were also aware of the cost of subjectivity and filtering by the interpreter, which would restrict the amount and perspectives of the information that participants receive. P10 who desires complete information and the freedom to decide on which information is important commented: *"[The sighted interpreter] edits all the information for me, and I keep pointing out [that] I want to edit the information. I want [the interpreter] to give me all the information; I'll make up the decision [on] what to omit."*

Participants also reported emotional and social implications, as sighted helpers might become frustrated when communication is not effective, leading to potential social costs. P10 continued: *"I'm married, and I want to stay married, so I only get my wife to help me so often, because we tend to fight early on in the stage of [her] interpretation."* P10 then added that when the emotion gets intense, giving up on accessing the diagram happens often as well: *"A lot of time you just give up because you don't want to alienate a person who's trying to help you. I mean it becomes an interpersonal dynamic - it can be challenging."* P8 also commented: *"You can sometimes drive people crazy when you really want to understand something, and some people don't have the ability to explain it to you easily. People get frustrated because they want you to understand with their explanation, and if you're not getting it, then there's a frustration level."*

Moreover, others' descriptions could be inconsistent. P8 had to sometimes combine multiple perspectives from others: *"Everybody has a different way of explaining something, and if somebody missed some key things, then it's important to get an explanation from 2 or 3 people, and then I can put things together better."*

In a professional setting, P2 highlighted: *"The problem is, what if you can't ask colleagues, or what if your promotions are dependent on you, looking like a person who's self-reliant? Now you look like you need more help than you really do."*

Others have expressed concerns with privacy having someone looking at a digital diagram on their devices. P3 said: *"I just kind of feel weird about other people looking at my phone, even if I'm not looking at anything ... It's my private thing. But certainly, if it was like something that I needed right now, and it was a time-sensitive thing, and there were people around to ask when I would ask them, then yes[, I would ask them]."*

A Lack of Reinforcement of Standards: Alt-text standards exist ([10, 24, 50]) and 7 participants reported advantages of having alt-text as a way of accessing diagrams outlined in Section 3.2.2.2. However, such standards have not been reinforced especially for digital diagrams. P2 said: *"I think there need to be standards that need to not only be put out there, but also enforced and monitored so that if there's something that's outdated the standards can be updated ... or made to fit more evolving needs ... They're not even taught mostly in lots of web development courses and universities these days."* This resonates with comments from P1 and P2 in Section 3.2.2.2 that despite existing standards for producing accessible information, many people are still not aware of them or do not know how to apply them effectively.

3.2.3 Unnecessary Diagrams

Diagrams are not always necessary because some of them can often be easily described in text. As explained by P2, people have the tendency to think that visuals are *"more fun"* and easier to understand than text, but this depends on the information and context. For example, P2 mentions that diagrams used in coursework for professional development packages *"can be so easily described in a couple of paragraphs"*. Additionally, certain information in a diagram may not be essential for the learning outcome and does not need to be included. In some cases, diagrams can be very simple and the textual version of them can be easily understood.

This also contributes to the difficulty of having equal access. P10 said: *"Often ... [a diagram is] not that important [and] I can do without knowing this information, which is too bad because you're basically not being able to access the same information of your [so-called] sighted counterparts, whether it's as personally or professionally, so you're acting on less information."*

3.2.4 Summary of Answers to Research Questions

Below we summarize the answers to the 4 RQs.

Answering RQ1: The majority of participants indicated that diagrammatic information is currently accessed through written or verbal descriptions, or tactile diagrams based on the representations of the diagram and the availability of tools such as screen readers. Encounters with diagrams usually happen in work, education, or personal settings.

Answering RQ2: Typical challenges with accessing diagrams from what participants reported include diagrams simply being inaccessible or not meeting the needs of the people accessing them. When the diagram gets too large or complex, it is hard to understand or takes time and effort to do so. Asking others can be inconvenient, not helpful, subjective, and frustrating, and it can also raise privacy concerns if someone else must look at the diagrams shown on personal devices.

Answering RQ3: Participants reported several current solutions to these challenges, including dividing information into chunks, gaining an overview before details, using different tactile textures or symbols, combining multiple perspectives or sources of information, and seeking external help or support. However, these strategies also had limitations or trade-offs, such as time cost, monetary cost, social cost, privacy cost, or cognitive load.

Answering RQ4: Some suggestions or ideas that could benefit future technology design include enforcing and monitoring accessibility standards (P2, 10, 15). For a complex diagram, breaking it down into smaller parts can make it easier to understand (P2). The components in a diagram should be nicely spaced-out reducing confusion caused by clutters (P3, 11). Simplicity and conciseness are important to only show what is relevant (P3, 4, 14). In terms of using audio to help access diagrams, distinct sounds should be used for different purposes (P1, 2). There needs to be a legend for the audio mappings between sounds and their meanings for people to refer back to (P1, 2). When navigating a diagram, there should be a mechanism to keep people on track (P11, 13, 15).

3.3 Insights

The analysis that we carried out on the interviews have led us to conceptualize the difficulties and challenges and behavior when accessing diagrams. In this section, we report a conceptualization of this space based on the findings above as well as other comments and ideas by participants that go beyond the answers to RQ 1-4.

3.3.1 A Framework to Incorporate the Findings: Ladder of Diagram Accessibility

We have conceptualized that there are 6 levels of diagram accessibility from the lowest to the highest degrees of accessibility: not knowing the existence of a diagram; knowing a placeholder of a diagram; accessing a static single view of a diagram; accessing multiple perspectives of a diagram; conducting dynamic queries about a diagram; and achieving complete and equal access to diagram data. We climb up the ladder and explain each level of it. As we explain, we also make connections to our findings and the data-information-knowledge-wisdom (DIKW) pyramid [81, 35].

Based on previous literature reviews [49, 63], the levels of information access from the ladder also partially align with the information granularity and accessibility dimensions, although these are meant for all data visualizations and not just diagrams. They have found that the most basic level of information access is informing about the existence of a visualization, which also aligns with the lower levels of our proposed ladder of diagram access. The next level is about overview, and the highest level is about gaining different types of details such as data values, data trends, and context. These also align with the information people could gather and interpret at the higher levels on the ladder. We also highlight the challenges preventing people from moving up the ladder at each level, opportunities enabling people to move up at each level, and the various pathways navigating through the levels.

3.3.1.1 Bottom Level (1): Not Knowing the Existence of a Diagram

At the bottom of the ladder, people are unable to perceive a diagram because its existence cannot be known. They have no options but miss the information involuntarily. P12 showcases an example where the screen reader skips a diagram completely, and P15 also commented *“I honestly wouldn’t be able to tell you if there was a diagram there or not.”* In a related example by P14, a diagram in a PowerPoint slide reads “embedded object” by the screen reader, signalling that there might be information, but the placeholder fails to inform P14 it is a diagram, leaving P14 confused and *“[had] no idea what that [meant].”*

3.3.1.2 Level 2: Knowing a Placeholder of a Diagram

Moving up a level, the existence of a diagram is made known to people with a placeholder. For example, screen readers can at least inform people that a “graphic” exists (but not all the time). P9 showcased this in our interview and commented *“[It] means nothing to me”* and this is not useful in accessing the actual diagrammatic information. Without such access, people cannot decide if the diagram is relevant to their needs and if its information should be accessed or to be ignored. P10 commented *“[I’m] constantly making the assessment [of] how worthwhile [the diagram is] for me to understand”* and this is confirmed by 4 other participants, but they cannot do so with just a placeholder.

This places an extra load on the people because they now have to spend more time and effort to find out what the diagram is about. This disadvantages people with visual impairments differentially because they do not have the same access to the visual information that sighted people have when they encounter a diagram. They may miss out on important or interesting information if they decide to skip a diagram that has only a placeholder.

Independent access to diagrams is also impossible. P1 shared the fear of getting a professional designation: *"I'm quite terrified about that because ... I don't know yet if they have an accessible way [for me] to access ... the study guide."* From here, people could try workarounds, or they just give up when no other resources exist for them, or the workarounds take too much time or effort. These will be explained in detail at the level above on the ladder because they also apply there.

3.3.1.3 Level 3: Static Single View of a Diagram

Most of the data that we gathered from the interviews is about this level of the ladder. This implies that people can often access a single, static view of a diagram, such as a textual, verbal, or tactile representation. However, their ability to understand the diagram to meet their needs depends on various factors (explained in Section 3.2.2). If these factors are not favorable, people may have negative experiences or be unable to comprehend a diagram successfully.

If the static single view of a diagram is limited, it can be difficult for people to understand. This can happen if the diagram is not produced with accessibility in mind, or if the diagram contains too much or complex information. This makes it insufficient, confusing, or difficult for people to understand, visualize or retain the information. As a result, independent access to the information becomes harder. P12 said: *"Sometimes there're things in diagrams that I don't know if they're there until somebody points it out."* Again, people might choose to give up, or turn to workarounds such as asking others who are sighted, finding alternative sources of information, using other tools, and repeating accessing the diagram in hope to gain more understanding of it. The associated challenges with asking others have been explained under results (Section 3.2).

Another common workaround is finding alternative sources for the same information elsewhere (N=11). P2 gave an example of not being able to access a hiking map and resorting to finding other *"people's experiences from their own written journals ... to get more details [to] fill in the blanks."* P9 also searched for alternative diagrams on search engines to find one with alternative text description if a sufficient degree of knowledge of the diagrammatic information was attained. However, there is no guarantee that the alternative sources are accessible and useful to the people, and finding a good one can be time-consuming. P10 said: *"When I don't get that information ... I go to the other avenues which that information is [in] an accessible format for me ... I read through all the information, so it's a much more cumbersome and time involved way of discerning the information. But at the end of the day, it's the same information, just I've had to spend a lot more energy to get it rather than looking at the diagram."* Additional tools could also help sometimes but they also have their own limitations. P5 commented on the usage of optical character recognition (OCR) to grab the text from a diagram going from left to right and from top to bottom, but it could also be time-consuming and sometimes limited in reflecting the connections between diagram entities. Lastly, people could try again accessing the same view of the diagram, but this takes efforts and time, is limited to one perspective only, and might create more frustrations. P3 said: *"I would try until the bitter end, until it looks like I can't handle this anymore, like I'm frustrated."* If people succeed in their workarounds and are able to gain more perspectives of the diagram beyond the single view given previously, they have advanced another level up the ladder.

Nevertheless, if the factors affecting access to a diagram are positive, people can have a useful single view of the diagram. This means that this diagram takes advantage of its representation (being digital, tactile, printed, or verbal), has good accessibility, and is clear to understand. It also means that the diagram meets the audience's goals. People can successfully complete multiple tasks at this level of the ladder and above. Such tasks include distinguishing different components of a diagram, finding relations, finding whether a certain element is in the diagram, and comparing one data element to another. Most of these are related to the data and information level in the DIKW pyramid, with one exception being data comparison which is more about knowledge.

We notice that a static one-size-fits-all type of solutions that is sometimes offered as the golden standard (e.g., a well-made alt-text description) does not optimally support people when they have different task needs or goals. There are tasks which people with visual impairments could perform to help themselves move up a level on the ladder by gaining more perspectives of the diagram, such as gaining an overview, and making notes. These new perspectives could help them to achieve more knowledge as well.

An overview of a diagram refers to a quick summary of the semantic information, the spatial information, or both. 14 participants commented that they would try to gain an overview first. P1 said: *"I would probably use [an overview] to search the information.*

I would definitely use it as a reference going forward, and to get the key information I need from it." P8 also said: *"If I had a basic outline of [the diagram], it would speed up the understanding tremendously."*

Note-taking is another important task that can be easily overlooked. Several participants mentioned the need to save information they have processed about a diagram for future reference or to enhance their memory. This can be done through making notes in separate documents, or annotating which involves inserting information directly into the diagram representation. Both methods help with understanding and memorizing the diagram. If given an editable written description of a diagram, P1 and P4 commented that they would utilize special symbols to be inserted into the description for future reference by searching for the symbols. If not editable, additional notes are often made. P1 said: *"I'll write out in text as best I can…noting all the pertinent information within it."*

3.3.1.4 Level 4: Static Multiple Perspectives of a Diagram

Moving up another level, having access to multiple perspectives of a diagram can provide people with a more comprehensive understanding of the information it contains. By viewing the diagram from different angles, people can gain different levels of information, cross-reference with other perspectives, and find commonalities between them. This helps with gaining knowledge about the diagram. Additionally, tasks such as summarization and reflection allow people to gain deeper insights into the information presented in the diagram.

Cross-referencing is a way to reinforce their understanding of a diagram, if multiple perspectives exist and can be accessed. P2 said: *"As you've seen in textbooks … the diagram is there to reinforce the learning outcome or to support learning. So then, what you can do is you can take the more clunky road of trying to figure it out just from the text and the writing in the book, or vice versa."*

Finding commonalities between perspectives or views can also be achieved. P2 commented: *"Looking at diagrams to set up tarps with ropes to make different kinds of shelters … Look for the commonalities … and then afterwards look more deeply into each setup to determine the specifics … You're looking for common words as well, for example … check which knot is used … [and] twenty-five different setups … may all use the same knot."*

Summarization of multiple perspectives is one way to generate insights. P14 explained using an example diagram of a theoretical model and the written article containing it: *"Summarizing this model [is basically] what the original article did, but in my own words*

... of explaining this model and then also discussing how this model is relevant to the study [for my] dissertation, so how it's relevant, how it applies."

This also leads to the task of reflection. P12 elaborates: *"Typically a diagram ... is meant to bring other information to mind, right? It's a concise guide to a bigger theory ... You might ask related questions ... and consider them and discuss. So ... what do you think this diagram means? And then starting to think about examples, or why is it important that these [things in the diagram] are related? How, why do they need to be together as opposed to separate?"*

Although this level on the ladder can enable better accessibility, it still has some barriers. A lack of consistency in the information accessed from multiple perspectives could happen, as outlined by P11: *"The challenge is ... [to] work out [multiple accesses to information] that are going to work universally [so that] I, the next person, and the next person after will be able to interpret them all in a similar way."* P10 also commented: *"It might work on one set of charts, [but] even within the same website it [might not] work on another."* More confusion could be resulted, and more time needs to be invested to understand the inconsistent differences between perspectives. Moreover, when switching between an overview and the details, information might get missed. P14 demonstrated that during the process of understanding the outline of the diagram as an overview, the details of entities were neglected: When P14 read the text of the diagram entities, connections between entities were sometimes missed. As we continue moving up the ladder of access, these barriers could be better overcome.

3.3.1.5 Level 5: Dynamic Queries about a Diagram

The higher we are on the ladder, the less data we have from the interviews implying that few people have achieved higher levels of accessibility. This level is about people being able to conduct dynamic queries about a diagram. For example, a tool P5 mentioned is an app called "Be My Eyes" [33] which has volunteers helping people see their surroundings through the camera view. In this way, a back-and-forth conversation between the volunteer and P5 can help P5 ask more targeted questions to gain a better understanding of the diagram. This can be more effective than asking a sighted person who might not have the expertise or patience to explain the diagram well, or who might not be available when needed. Moreover, a back-and-forth interrogation of the diagram with someone who is sighted can result in social cost and negative emotions (as explained in Section 3.2.2.3) as the helper can become alienated, frustrated, or impatient when they cannot explain the diagram in the way the inquirer wants. Asking other people for help can also depend on the availability and willingness of the sighted person, which can limit the access to diagrams at any time or place.

P3 also commented on the experience of encountering a diagram on Twitter: *"I would probably just tweet the person and ... ask a specific question about it."* P8 also described a dialogue with architects to clarify the exact structure of a building. Through such interactive exchange of information, people could more effectively gather and filter information, and form knowledge and insights related to want they want to know.

3.3.1.6 Top Level (6): Complete (and Equal) Access to a Diagram

The complete access to a diagram would help people independently consume a diagram. This also helps achieve equal access where all individuals can *"have the same opportunity to get an equal level of understanding"* of a diagram, as defined by P2, but no data suggests that this has been achieved. Going beyond this, it is also challenging to achieve equal access without costing time or money. Even when complete access to a diagram is granted, more time could be spent on absorbing all the information. P14 commented: *"I think a person with sight really could like, boom, they could kind of answer the simple question really quite quickly, just like looking visually at the bars ... For me, it took more time and effort."*

We also note that a diagram is just a way to represent information, and this top level aims to achieve complete access to the information (and not necessarily the visualization of a diagram). This means that the top level of diagram access is not just about being able to completely see or understand the diagrammatic representation itself, but also about being able to access the underlying information that the diagram represents. The goal is to provide equal access to the information, regardless of how it is presented or represented. This can empower people with visual impairments to make their own decisions and draw their own insights from the data, without being limited by the format or presentation of the information.

3.3.2 More on the Relationship Between the DKIW Pyramid and the Ladder of Diagram Accessibility

Being on different levels of access on the ladder affect which level people could be at in the DKIW pyramid as well, and generally a higher degree of accessibility leads to a higher-level representation of knowledge, but sometimes, a higher level in the DKIW pyramid can be achieved at a lower level on the ladder of access. For example, a well-constructed description of a bar chart is sufficient to not only read people the values as data and information but also support them in comparing the values as knowledge, as P4 confirmed. In addition, moving up the DKIW pyramid and generating insights are not what people want to achieve all the time because these depend on the goals of the individual.

It is important to highlight that because the accessible information tends to be a processed filter of the diagram data, such as an interpreted description, people might not get a chance to access the complete objective data. The accessible view(s) of a diagram could already contain information, knowledge, and even insights which people could directly gather, but this limits their freedom for them to decide on what to do with the data themselves. The top level on the ladder would provide the complete information without any filtering or interpretation, allowing the user to access the diagram as it is. People at the top of the ladder could also choose how to filter or interpret the information according to their own needs and goals, rather than relying on someone else's perspective.

Chapter 4

Designing a System Enabling Touch-and-Audio-based Diagram Access: TADA

The formative study in Chapter 3 provides valuable results and insights into how technology can improve the accessibility of diagrams. These insights, combined with existing literature, have informed the design of our Touch-and-Audio-based Diagram Access (TADA) system, an application specifically created to enhance diagram accessibility. The main goals of the design are outlined in Section 4.1. The approach for designing the tool is then codified into a series of design principles, which are detailed in Section 4.1.1. The design of the system, presented in Section 4.2, incorporates these design principles through carefully considered design choices from Section 4.1.2. This ensures that TADA effectively addresses the challenges of diagram accessibility and provides a user-friendly solution.

4.1 Design Considerations

We aim for two main design goals (DGs):

DG1 : Help people move up the ladder of diagram accessibility (from Section 3.3.1) to achieve higher levels of access.

DG2 : Reduce barriers to accessing the designed system. To achieve this goal, the design should strive to maintain a balance between leveraging the existing knowledge of users and introducing new concepts, ensuring a manageable learning curve. Familiar concepts should be retained where possible, and new solutions should only be introduced where no current solutions exist.

4.1.1 Design Principles

To achieve our goals, we encoded a series of design principles (DPs) that reflect our approach to designing the tool. These principles, informed by both the results of our previous formative study, and by experience derived from previous work, provide a framework for making consistent design decisions. While other design approaches are possible, we believe that our principles offer a strong foundation for an effective tool to improve diagram accessibility.

DP1. Distinguish different diagram information types clearly: A typical node-link diagram consists of nodes and links which might contain embedded labels or properties. For example, a transportation connection diagram could have each node as a station with its name as one property. Different nodes, links, labels, and other potential information types need to be clearly distinguishable to avoid confusion.

DP2: Rely on Spatiality: With DP2, the system portrays the spatial locations of different diagram components to help people understand where things are, associate an object with its spatial information, and potentially form a mental map of a diagram. In some diagrams, spatial information is also inherently important to understanding the diagram especially when directions and navigation are involved. The use of spatial information also supports DP1.

DP3: Support Multiple Levels of Information Access: To understand a diagram effectively, multiple perspectives of it such as different overviews and details could be provided. People should be able to choose which of them to be accessed first preferably as an easier starting point, and to cross-reference information between the different perspectives to help reinforce their understanding. When switching between different levels, the consistency of low-level information needs to be preserved to avoid confusion. For example, the representation of a diagram component is the same at all levels of access.

DP4: Provide Multiple Interactions to Access the Same Type of Information: The system should provide the flexibility for people to choose from multiple supported interaction techniques or even combine them to access the same type of information.

This allows users to tailor their experience and interact with the system in a way that best suits their needs and preferences. It should not restrict users to using only one single interaction technique to acquire a piece of information.

DP5: Support Parallel Interactive Access: DP5 would allow people to have the option to conduct different interaction techniques simultaneously. Parallel interactive access can potentially increase the efficiency of access and better support someone's primary purpose of accessing a diagram.

DP6: Re-think, Not Translate: While a direct translation from one modality to another to improve accessibility could work, there is no guarantee that this would be optimized to help people achieve their goals when accessing a diagram. A paradigm shift here is important for making effective design decisions for non-visual access to support people with visual impairments as much as we can.

4.1.2 High Level Design Choices

In this section, we discuss the high level design choices of TADA with references to the DPs mentioned in the section above.

A practical way to improve accessibility is to minimize the associated monetary cost as mentioned by multiple participants from the interview study. We decide that we want to rely on a commercial touchscreen device to be the basis of our prototype due to its wide availability. As participants from the formative study suggested (Section 3.2.2.3), external tools can be expensive and we want to make it as affordable as possible. Furthermore, we prefer a tablet device with a bigger screen than a phone due to its bigger space for people to interact with a diagram on the touchscreen, reinforcing the spatiality of the diagram components (DP2). The advantage of using fingers to understand spatiality is the use of haptic memory. While the touchscreen is just a flat surface without any tactile feedback to touches, proprioception (which is the sense of body position and movement) could help people recall the spatial information of diagram components [96] by using their fingers to explore the diagram (DP2).

With lessons from how diagrams are currently made accessible, we learned that a diagram should be spaced-out to ensure clarity as people explore it (DP1, 2). In the case of spatial information from the data not being available, the force-directed layout [36] of a node-link diagram can be utilized which ensures nodes repel each other forming varying degrees of space between them, and naturally forms clusters of nodes which could help improve readability [66].

We decided that TADA should support access to multiple levels of information (DP3). Informed by the previous literature reviews on accessible visualizations [63, 49], task taxonomy of node-link diagrams [6, 68], and the interview study findings, we include gaining overview, gaining details, searching, navigating, filtering, and alt-text description. An overview is important to gather a high-level idea or summary of the diagram, and details inform more about the diagram components and structure. Searching allows people to look for where things are based on what they look for specifically. In scenarios where understanding the connections between nodes is important, navigation informs people about how to go from point A to point B. Filtering for only certain type(s) of nodes and their connections would help reduce the complexity of a diagram. Lastly, alt-text is still included as a common way to gather information on a diagram. To support access to these levels of information, we utilize gestures and speech commands as input to the system, and audio feedback (including non-speech sounds and speech) as output from the system. They together form the interaction techniques that TADA supports, which will be explained in detail in the next section on the design of TADA. Although different interaction techniques are meant for different purposes, some of them allow people to access the same type of diagram information (DP4). For example, all interactions inform spatial information of diagram components. There is also information that stays consistent using different interaction techniques, like the audio used to portray nodes and links.

Audio is an alternative way for a typical tablet to output information other than visuals (haptic feedback technology, which provides tactile sensations to users, is currently only available on a limited number of high-end tablet devices). We therefore decided to use audio and touch to be the primary modalities of access. The audio for different types and levels of information needs to be distinct and informative to avoid confusion (DP1). We make use of the following dimensions of audio design: timbre, pitch, volume, rhythm, continuous or discrete, and sound or speech. We use two sounds with vastly different timbres for nodes and links. They are pitched differently based on the number of neighbours a node has, and the length of a link. When a finger stays on a node or link, its sound plays continuously until the finger leaves. If the finger sweeps quickly, the sounds of nodes and links become shorter and more discrete. A different timbre is used for the system to guide certain actions performed by people continuously, and its varying volume informs the proximity of the finger toward a target position. Another different timbre is used to highlight diagram components which are filtered out if they are as an additional sound layered on top of existing audio. All the above audio are non-speech sounds. Speech is used for searching to inform directions, and for textual details. There is an audio legend explaining what different sounds mean to remind people, and alt-text for people to gather a description of the diagram. Rhythm is also used for some interactions to reinforce certain information such as connectivity and proximity. We did not decide to involve chords or other ways to present multiple pieces of information simultaneously using pitch because the audio could become more musically complex for people to interpret and we want to maximize the other more differentiating audio dimensions first. The combinations of interaction and audio designs are examples of how we re-think non-visual access to diagrams, instead of directly translating a diagram into an audio form (DP6).

We decided TADA should support multi-touch because it opens up the possibility of more gesture designs and supports two-handed interactions. This means two things. One, a multi-touch gesture could be performed either using one hand or two hands depending on preference. Two, it is possible to perform two tasks with separate hands on the same screen to access different information in parallel (DP3, 5). To achieve this, we utilize the concept of locality to distinguish what each hand is trying to do without confusing the two. Generally, a gesture is only invoked if its trigger fingers are within proximity to each other in a local area. As such, another gesture performed by the other hand at another part of the screen would be recognized differently. For example, one can choose to keep a finger as a reference point and use the other hand to perform other gestures to access different pieces of information. The use of locality also avoids accidental taps because of the condition of a dwelling finger nearby, aligning with one of the accessible touchscreen guidelines from McGookin et al. [71] to prevent accidental gestures with a short impact.

4.2 System Design

This section details the interaction design of TADA. The DPs from Section 4.1.1 guided design of the interaction techniques and features of the tool and are naturally constrained by the top level choices that we describe in Section 4.1.2 indicated with references to the DPs.

TADA reads diagram data from GraphML file(s) [16] and represents the data digitally in a perceptible way with auditory and visual cues. People can interact with diagrams using touch, gestures, and speech as input, and the system generates audio in real time as output. Diagram elements such as nodes and links produce distinct sounds when people interact with them. Various interaction techniques (Figure 4.1) are designed to support user tasks such as overview, gaining details, finding connected links, navigating, searching, and filtering. We explain these in detail below.

4.2.1 Audio Design

We briefly explain the mappings between diagram information, interaction techniques, and the non-speech audio feedback below:

- Nodes produce a french-horn sound, and links produce a plucked guitar sound.
- A higher-pitched node means it has more connections than other nodes; A higher-pitched link means it is shorter than other links.
  - The pitch frequency (in Hz) of a node is calculated by:

    $$f_{node} = \text{number of connections} \times 20 + 50$$

  - The frequency (in Hz) of a link is calculated by:

    $$f_{link} = \frac{1}{\text{length in cm}} \times 1215 + 50$$

- A sustained french-horn or guitar sound indicates dwelling on a node or link respectively; A discrete sound indicates sweeping over a node or link.
- A continuous pure tone indicates the proximity to a diagram element. The pitch frequency is 150 Hz.

The variation in sound frequencies (pitch) is adequate for now, but it would have to be configurable or adapted for other diagrams with more variability. For example, if a diagram has too many nodes or links with different degrees or lengths, the pitch range might not be enough to distinguish them clearly.

4.2.2 Supported Interaction Techniques

This section explains how the design of interaction techniques (shown in Figure 4.1) in TADA.

1. Single-finger Sweep: People can conduct the single-finger sweep to learn about the general overview of connectivity and distribution of nodes across a diagram. When a finger touches a node, the system produces a french-horn sound; When a finger crosses a link, the system produces a plucked guitar sound. A higher-pitched node means it has more connections than other nodes; A higher-pitched link means it is shorter than other links. As the finger sweeps around quickly, the audio of a node or link is short and discrete as the finger sweeps by it. In an area which is dense with nodes, the audio sounds like a grouping of short bursts of french-horn sounds; In an area which is dense with connections, the audio sounds like a cluster of short guitar strings plucked together.

If a finger stays on a node or link, the sound sustains and fades to a fainter volume until the finger leaves to indicate the dwelling status. Once a finger enters a node area, the area expands to make it harder for the finger to leave. This helps with the potential issue of the dwelling finger drifting away from the node position because finger drifting might result in knowing inaccurate spatial information which could confuse people's perception.

The speed of movement of the finger matters and affects the discreteness of the sound. As the finger slows down, people can explore the audio in more detail interpreting pitches more instead of the number of discrete sounds produced together.

We note that although this interaction involves a single finger, multiple fingers are still supported. This is explained more below in the next Section, For example, if two fingers each dwell on a different node, we hear a combination of the two french-horn sounds played together. One can also choose to lift a finger up while keeping the other down and switch between them repeatedly to conduct a compare-and-contrast operation.

2. Five-finger Dome: The five-finger dome allows people to gain an overview of only a sub-area within a diagram. By shaping a hand into a dome shape with fingers pointing downwards and placing it onto the screen, an area is formed underneath the hand between the fingers. An audio stream starts playing, and it only involves the nodes and their connected links within this area under the five-finger dome. It begins with a "ding" sound, followed by the first node's french-horn sound, the first node's connection guitar sound(s), the second node's sound, the second node's connection sounds, and so on until all nodes and their connection sounds within the area are played. Then, another "ding" sound plays, and the same audio stream repeats infinitely as long as the hand is on the screen. The sequence of nodes follows their spatial positions, starting from the center of the dome and going outwards.

The rhythm of the node and link sounds reinforces the characteristics of connectivity within the dome area in addition to the pitches of the nodes. Each loop of the audio stream takes a set amount of seconds. This duration is divided by the number of nodes covered by the dome into even intervals. Therefore, if one hears a faster rhythm of french-horn sounds played, the dome area currently has more nodes than other areas explored. For each interval, it is further divided into two halves. The first half of each interval is for the node; The second half is for its connected links. Therefore, if one hears faster pacing of links played for a node, this node is more connected than other nodes in the same area. Several examples can be used to illustrate the relationship between the audio summary and the structure of the dome area. If there are multiple quick node sounds but few link sounds, this indicates that the dome area contains many nodes with relatively few connections. Conversely, if there are few node sounds but each is accompanied by quick link sounds, this suggests that there are dense connections between nodes, but the overall number of nodes is limited. Finally, if each node sound is followed by link sounds with roughly the same pacing, this implies that the nodes generally have an equal number of connections.

People also have the flexibility to dynamically reposition their dome, or expand and shrink the size of the area of the dome. Any change in the set of nodes covered by the dome would trigger a new "ding" followed by the adjusted audio sequence. As such, people can gather mini summaries of sub-areas of a diagram as they explore using the fiver-finger dome.

3. Dwell + Tap: All interaction techniques that involve a "+" in their name refer to the use of two simultaneous actions with different touches. The notion of locality applies to such interaction techniques, meaning that the second action can only be triggered within the proximal area of the first action. This is in line with one of the recommendations for creating non-visual graphics [89], which suggests taking advantage of multi-touch and using other fingers as reference points. It also considers the findings from a previous study [71] which showed accidental taps should be avoided, and the locality design helps prevent this.

Dwell + tap is a way for people to gain more details about the embedded or annotated information of a node or link which can have any number of labels or attributes. The first finger dwells on a node or link, and the second finger taps next to the first finger. Each tap triggers the system's speech synthesis engine to say a piece of information related to the node or link the first finger is on. When the user has played over all the information, the next tap triggers the first piece of information again and subsequent taps continue the cycle. For example, if a node represents a person, then dwell + tap at this node would first say “Name: Alex”, and the second tap might say “Job: Engineer”. If no more information is available, the next tap would go back to the name.

Due to the locality design, if one chooses to put two fingers separately on two different diagram elements, taps associated with the closest dwelling finger would invoke the corresponding information of the closest elements dwelt on accordingly. This allows simultaneous exploration of the data of multiple diagram elements with separate hands. If multiple locality areas overlap with each other, the dwelling finger and its nearest second finger are considered to be a pair.

4. Dwell + Circle: With dwell + circle, people can find the links connected to a node. The first finger dwells on a node, and the second finger lands near the first finger. Without lifting, the second finger then either holds or starts moving in a circle around the first finger to trigger a continuous pure tone which informs about the proximity to the nearest connected link. This proximity guidance audio has varying volumes, where a louder volume means the second finger is closer to a link. The angle bisector between two consecutive links is silent.

There are now three types of continuous audio that play at the same time: the french-horn node sound as the first finger stays on the node; the guitar link sound as the second finger finds a connected link; and the proximity guidance sound. The system pauses the node sound to prioritize the proximity guidance and link audio at the moment. (However, if the second finger gets lifted off, the node audio will play again.) We paused the node sound because when all three play together, they can sometimes interfere with each other. As people rotate their fingers, they can still hear the link guitar sounds as the second finger crosses them. The relative positioning of the two fingers could help inform the direction of links, and the pitches of the link sounds could inform the lengths. If using two hands, we recommend people position the second finger below the first finger as the initial starting point of the circle to better ensure the completion of a circle to fully explore connected links.

5. Dwell + Radiate: Dwell + radiate is the natural follow-up gesture to dwell + circle once people have found which link they want to follow, to navigate from a node to its neighbour. During dwell + circle, if one is interested in a link and wants to follow it, the second finger moves toward the direction of the link from the node. The system then recognizes the dwell + radiate gesture and speaks “Following”, signifying the second finger is now following the link. Other links than the one being followed are muted to avoid possible confusions. The proximity guidance audio now indicates the distance to the target neighbouring node, rather than the distance to the nearest link. However, it still provides indirect information about the proximity to the nearest link. As the second finger gets closer to the target node, the volume of the proximity guidance audio increases. If it deviates away from the link, not only the sustained guitar sound would stop playing, but the proximity guidance audio also decreases in volume.

If someone is lost, they can wiggle their radiating finger a bit perpendicularly to the radial direction until they find the link again. This aligns with one of the guidelines on tracing a line [89] to take advantage of tracing strategies like zigzagging on a line.

If the radiating finger has followed the link and reaches the connected neighbouring node, the french-horn sound of it plays but with a fanfare twist signifying victory, by playing a short, initial horn burst followed by the original sustained horn sound. Now, the original second finger acts as the finger dwelling, and the same interactions can still continue navigating further.

6. Single-finger Flick-down for Speech Commands: A single-finger flick downward triggers speech recognition. The system then says "Listening" and listens to speech input by people. It expects the speech input to specify the type of command.

6.1. Speech Command for Searching, and Single-finger Follow. After the single-finger flick-down, people can say "Searching for [information to look for]" invokes search mode. The system then auto-recognizes the end of the speech, repeats the speech command for people to know what is recognized, and begins processing.

After processing, the system responds verbally with the results of the search. For example it could say "Nothing found", or "Found [number of results] result(s)" depending on the search. People can now start a search for the results by putting down a single finger onto the screen and holding. The system then starts to provide directions (left, right, bottom, up) verbally that direct the user's finger to the location of the nearest result. For example, if the result of the search for is in the lower left direction of the finger, the system would respond with "down, left" and loop this. The directions are dynamic to changes in the finger's position. If the finger moves diagonally, the directional instruction combines two axes together. Alternatively, the finger can also move in one axis at a time, and only directions in this axis are spoken to help people to find the point on the same axis as the target location. Moreover, we varied the pacing of the speech to inform about the remaining distance to the target. A slower pacing means the finger is closer. This is because as the pacing slows down, one would also slow down the finger movement to wait for the next piece of directional instruction to more precisely pinpoint the target location.

Once one axis of the target location is found, the system would say “stop”. Then, only the directions on the same axis as the target location are informed until the finger reaches the target. Similar to successful navigation, when a node is successfully found, the same fanfare-like horn sound is heard.

According to a study by Ramôa et al. [80], voice-based guidance is the most efficient method for helping people pinpoint a location, and it does not require prior training. We took this finding into account and additionally modified the speech speed based on the distance to target. We also introduced the diagonal or axial finger movements.

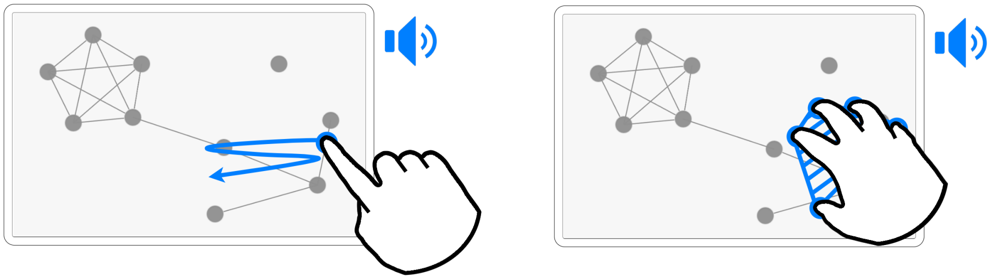
6.2. Speech Command for Filtering, and Interact as Per Usual. The initial trigger is still single-finger flick-down with the difference in the following speech command: “Filter by [attribute, value]” to initialize the filter mode. For example, after the single-finger flick-down, the system says “Listening”. People can then say “Filter by background, engineering”. The system recognizes the end of the speech, repeats the speech, and enters into filtering mode. All interaction techniques still apply.

The filtering mode is introduced as a way to reduce the complexity of a diagram by specifying an attribute value to be filtered for, and only nodes with this value for the attribute and their connections are best preserved. The attributes and values that users can filter for are the same information accessible through the dwell + tap interaction. Once a person has applied a filter based on a specific value, such as a background of engineering, only the nodes that meet the filter criteria and their connections are fully accessible. To preserve some degree of awareness of the filtered-out elements, these nodes and their connections still produce sounds, but at a fainter volume and with a static white noise played on top of them. This signifies that these elements have been filtered out and are not currently the focus of the person’s attention.

7. Single-finger Flick-right: People can listen to the alternative text description of the diagram by flicking a single finger to the right.

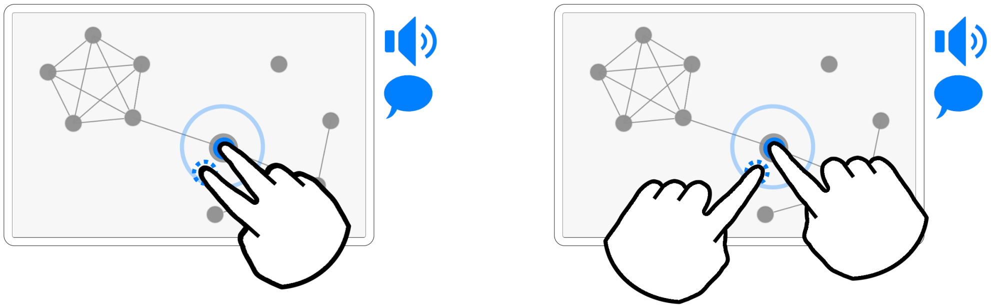
8. Double-finger Flick-right: A person can listen to the audio legend by flicking double fingers to the left. The audio legend is a description of the audio mappings, playing each type of non-speech audio feedback and describing what it means in speech.

9. Single-finger Flick-up: A single-finger flicks up cancels or reset the system behaviours triggered by other interaction techniques with speech feedback specifying what is canceled. For example, when performed right after double-finger flick-right, the audio legend stops playing, and the system says “Audio legend stopped”.



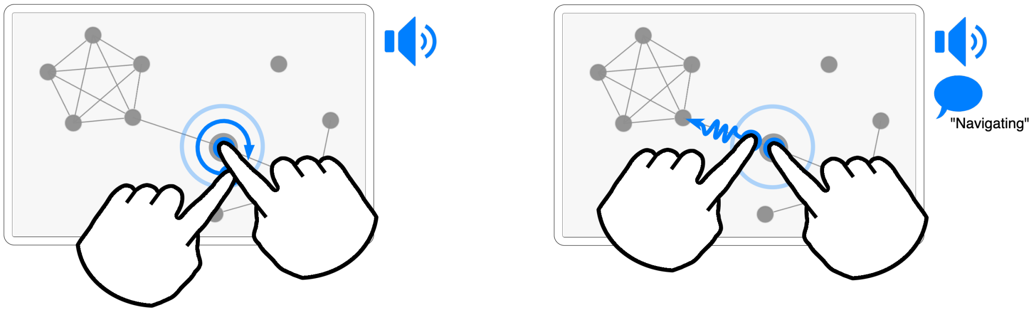
(a) Single-finger sweep

(b) Five-finger dome



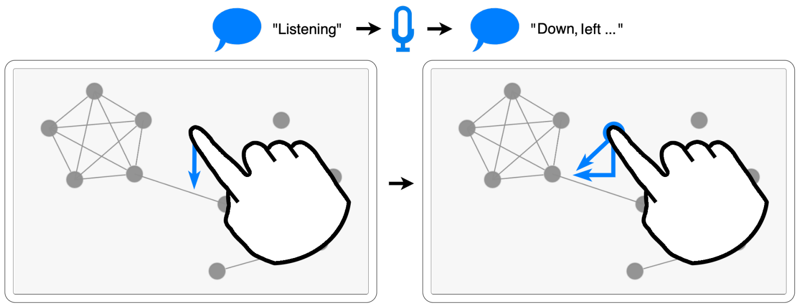
(c) Dwell + tap with one hand

(d) Dwell + tap with two hands



(e) Dwell + circle

(f) Dwell + radiate



(g) Single-finger flick-down, speech command for searching, and single-finger follow

Figure 4.1: Interaction techniques supported in TADA

Chapter 5

Implementing TADA

The TADA system is implemented using Unity [92], ChucK [93], and Chunity [11]. Unity is a game engine and C# scripts were programmed to parse diagram data, create the user interface and interactive diagram elements, recognize different gestures and their interactions with diagrams, and trigger the corresponding audio. Unity supports multiple platforms meaning that the system can run on tablets supporting various operating systems [88]. ChucK is a programming language that produces concurrent, strongly-timed procedural audio, and is responsible for the audio design and generation in the system. Chunity combines both ChucK and Unity and provides the technical communication bridge between the two. Native speech recognition using speech engines running either on iOS / iPadOS or Android is made possible with open-source codes [48].

The system includes several modules (Figure 5.1) that work together to make diagrams accessible through touch interactions, speech input, and audio feedback. The Diagram Module is responsible for parsing diagram data and creating the diagram elements in the interface. The Multi-touch Module recognizes touch gestures and their collisions with the diagram elements. The Controller Modules process and communicate the associated diagram information, and trigger audio feedback. The Audio Modules handle the audio design and production triggered by the different recognized interactions, and there is a speech engine responsible for text-to-speech and speech-to-text.

5.1 Diagram Module

The Diagram Module reads diagram data from GraphML [16] files. It has methods to parse, display, and layout the nodes and links of the graph. It uses dictionaries and lists as data structures to store the node and link data and types, and game objects and

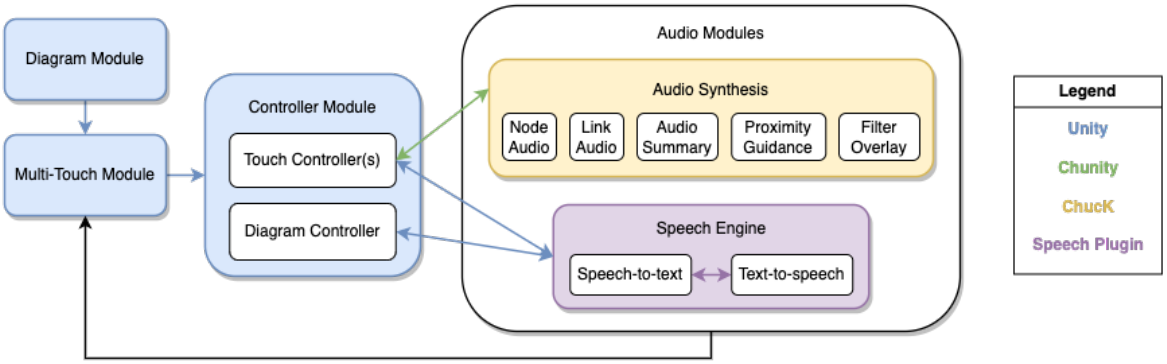


Figure 5.1: System architecture of TADA

associated scripts to create and manipulate them on the screen in Unity. For diagram data without spatial information, it uses the force-directed layout algorithm [36] to adjust the positions of the nodes and links based on their distances and connections. It also allows filtering the graph by node attribute values.

5.2 Multi-touch Module

The Multi-touch Module handles touch gestures, classifies them into different interaction modes, and communicates the information to the Controller Module for more logical processing. It recognizes various gestures based on the number, duration, direction and distance of the touches. It also keeps track of the diagram nodes or links which the fingers interacted with. There are two important threshold values, TIME_THRESHOLD and DIST_THRESHOLD, responsible for fine-tuning gesture recognition. TIME_THRESHOLD determines if the time duration between the start and end times of a touch or the time duration between two consecutive touches is within a certain range. DIST_THRESHOLD determines if the distance between two touches or the distance between the start and end positions of a touch point is within a certain range. We explain how gesture recognition works below.

Dwell or Sweep: The Multi-touch Module assumes this is the default gesture, unless the following different gestures are recognized.

Five-finger Dwell or Dome: The module checks if there are 5 touches on the screen that started within a short time duration of each other. If there are, it considers it as

Five-finger Dome gesture and creates a pentagon area that covers all 5 touches. This area is present as long as all 5 touches are present on the screen.

Two-finger Interactions (Dwell + Tap / Circle / Radiate): These three gestures have the condition that there has to be an existing touch dwelling on a diagram element (node or link) within the local area. The module checks the distance between the second-finger touch and other existing touches. If the distance is less than DIST_THRESHOLD, the method considers the second-finger touch and its closest existing touch to be a locality pair. The following all need to fulfill this locality condition first.

Tap: When a touch is ending in its phase, the Multi-touch Module checks the time duration between the start and end times of a touch to determine if it is a tap gesture. If the time duration is less than the TIME_THRESHOLD, the module considers it as a tap.

Circling: The module calculates the angle difference between the previous and current positions of two touches to determine if it is a circling gesture. If the angle difference is not zero, the module considers it as a circling gesture. Alternatively, if the second finger also dwells, the module also recognizes this to be circling.

Radiating: The module calculates the angle and distance differences between the previous and current positions of the two touches to determine if it is a radiating gesture as the touches move. If the distance difference is increasing, the angle difference is less than a certain threshold, and the distance between two touches is greater than or equal to the product of DIST_THRESHOLD and 0.65, the module considers it as a radiating gesture.

Flick: The module checks the distance and time duration between the start and end positions of a touch point to determine if it is a flick gesture. If the distance travelled by the touch in either the horizontal or vertical direction is greater than $\frac{\text{DIST\_THRESHOLD}}{1.5}$ and the time duration of the touch is less than TIME_THRESHOLD, the module considers it as a flick gesture. The module then calculates the direction of the flick based on the sign of the difference between the start and end positions of the touch in both the horizontal or vertical directions. The module also checks for the number of touches flicking. Depending on the direction and the number of fingers used for searching and flicking, different interaction modes such as filtering, reading alt-text, and reading audio-legend are identified.

There exists a separate data structure, Touch Info, that stores information about each touch, such as touch ID, start time, start position, interaction mode detected, and the associated diagram elements if they are interacted with. There is also a set of technical mappings assigning the interaction modes to each finger touch to communicate this information between modules.

5.2.1 Fine-tuning the Threshold Values with Informal Pilot Studies

We conducted informal pilot studies with three people who are sighted and had their eyes closed to help refine some details of the system design, particularly fine-tuning the threshold values (TIME_THRESHOLD and DIST_THRESHOLD). The recognition of tapping is now more tolerable for people, increasing the amount of time a finger stays on a screen before it is recognized as a hold. The locality area detection is also improved, making it bigger for people with bigger hands. The five-finger recognition is enhanced, increasing the duration between the landing of each finger, and making it more tolerable.

5.3 Controller Modules

There are two major kinds of Controller Module: Touch Controller(s) and Diagram Controller. Each new touch input is dynamically assigned a Touch Controller. The Multi-touch Module assigns each touch with particular interaction modes, and the Touch Controller for each touch input can understand what interaction the finger is doing and provides the respective system actions. Additionally, there is a sub-type of Touch Controller, the Dome Touch Controller, for the five-finger dome interaction specifically to handle the logical processing with the dome area. The Diagram Controller is for controlling the whole diagram instead of the touch input(s).

5.3.1 Touch Controller(s)

Each Touch Controller serves as the mediator between the Multi-touch Module and the Audio Modules by handling the logic for different touch interactions with diagram elements such as nodes and links. It also triggers audio synthesis and speech feedback based on the interaction modes and collisions. It gathers information such as the number of neighbours a node has and the length of a link and sends these to the Audio Modules to inform the pitch. For the five-finger dome, the Dome Touch Controller gathers these pieces of information from all the nodes and links interacted with the dome area, compiles it, and sends it to the Audio Modules.

For the dwelling + tap interaction, the Touch Controller of the dwelling finger is responsible for the diagram element dwelt on. The Touch Controller of the tapping finger sends pieces of the embedded textual data from the diagram element to the speech engine, and tracks which piece of information was last read, so that sequential taps can also explore the information sequentially.

The dwell + circle interaction lets the user hear the proximity guidance audio of the neighbouring nodes by circling around a node. The Touch Controller of the circling finger measures the angle between the dwelt node and the finger, and compares it with the angles of the connected links. It then adjusts the volume of the proximity guidance audio based on how close the circling finger is to a connected link. By using a logarithmic scale for the volume, the volume changes more drastically at the two ends of the scale.

For the dwell + radiate interaction, the Touch Controller of the radiating finger calculates the distance between the target node and radiating finger, as well as the distance between the target and source nodes. It then calculates a distance factor based on these values dividing the former by the latter. This value is used to modulate the volume of the proximity guidance audio, and a logarithmic scale is used here as well. It also checks if the touch has arrived at the target node to trigger a triumph-like node sound.

The single-finger flick-down interaction activates speech recognition. The Touch Controller of the flicking finger triggers the Speech Engine to say “Listening” and wait for a voice command.

If the speech command is for searching, the Touch Controller starts by getting the node(s) that match the search query. If multiple results are found, it then loops through these nodes, calculates their distance from the finger position, and picks the closest. It also calculates the direction based on the angle of the node relative to the finger. Auditory feedback is then provided to the user about the location of the nodes that match the search query where the directions inform the speech instruction given back to the user, and the distance to the target affects the pacing of the speech instructions. The controller also checks if the touch has collided with any of the nodes that match the search query. If so, it plays the appropriate triumph-like audio and updates the state of the system accordingly. Otherwise, the system stays in searching mode.

If the speech command is for filtering, the Diagram Controller takes over because this is more to do with how diagram data is represented. This also holds true for the single-finger flick-right interaction.

The double-finger flick-right interaction activates the audio legend description. The Touch Controller triggers various types of non-speech audio feedback supported by TADA. After each audio type is played, it sends a description of what the audio means to the Speech Engine. The next audio type then plays.

The single-finger flick-up interaction quits any ongoing system operation, such as searching, filtering, or reading alt-text or audio legend. The Touch Controller checks which mode the system is in and sends a quit signal to stop it.

5.3.2 Diagram Controller

After initializing the filtering mode, the Diagram Controller triggers speech recognition to recognize what the user wants to filter for. After the user provides the speech command, it then receives the instructions on which nodes in the diagram to filter, and tags these nodes and their connected links to be in a filter state, which will, in turn, inform the Audio Modules to produce the appropriate audio for them.

To read the alt-text of a diagram, the Diagram Controller simply extracts the embedded alt-text from the diagram and sends it to the Speech Engine. If no alt-text exists, it then asks the Speech Engine to say “No alt-text found”.

5.4 Audio Modules

The Audio Modules consist of two parts: the ChucK programs responsible for audio synthesis, and the Speech Engine as a Unity plugin responsible for speech recognition and synthesis.

5.4.1 ChucK Programs

With ChucK, the audio synthesis is achieved in real-time and is dynamic based on the outputs from the Touch Controller(s). It uses an event-driven approach, where the outputs from the Touch Controller(s) trigger events. ChucK uses a concept called “shreds,” which are similar to threads in an operating system, to handle these events. When an event is triggered, a shred is spawned to produce audio based on the event type.

There are five ChucK Programs responsible for the audio synthesis of the nodes, links, audio summary (from the five-finger dome interaction), proximity guidance, and filtered audio overlay. Each node audio is a french-horn sound, and each link is a guitar string. The instrumental audio is produced by the Synthesis ToolKit [27, 83] included within ChucK [1]. The audio summary is constructed by a sampled “ding” sound, followed by the audio of given nodes and their connected links with duration calculated based on the description in the System Design Section, 4.2.1. The proximity guidance uses a sine wave oscillator which sounds like a pure tone. The filtered audio overlay dims the volumes of filtered nodes and links by half, and then adds a noise audio on top of them to signify these are filtered out.

5.4.2 Speech Engine

The Speech Engine uses external plugins [48] in Unity to utilize the text-to-speech and speech recognition functionalities natively supported by the mobile platforms, Android and iOS / iPadOS. The text-to-speech reads out designed phrases or instructions from the system to invoke user actions. It is customizable in terms of the pitch and rate of speech synthesis. The speech-to-text function is responsible for recognizing people’s responses which serve as instructions to invoke further system actions. It recognizes a stop in a spoken sentence from someone, automatically stops listening, and starts recognizing.

Chapter 6

Evaluating TADA

To evaluate the effectiveness of the Touch-and-Audio-based Diagram Access (TADA) system, we conducted a system evaluation study with 25 people with visual impairments. Our research questions (RQs) for the study included:

RQ1: To what degree are people with visual impairments successful in performing the experimental tasks utilizing the interaction techniques and understanding the audio information?

RQ2: How do people with visual impairments perceive the workload of using the system?

6.1 Methods

To answer the RQs, e designed experimental tasks to be performed using the supported interaction techniques, questions to ask about each task, and their evaluation metrics to evaluate the effectiveness of our designed system. We also observed the participants using the system, and included general questions to ask about what people think about each interaction technique. Additionally, we used NASA Task Load Index questionnaire (NASA-TLX [41, 40]) to ask about their perceived workload and opinions of the ease of use and difficulty they experienced while using the system. We verbally administered the demographic and NASA-TLX questionnaires to avoid document accessibility issues.

6.1.1 Participants

The participants ranged in age from 25 to 101, with a mean of 51.92, median of 52, and standard deviation (SD) of 16.35. Fourteen participants self identified as female, 10 as male, and 1 as transgender. In our study, although all participants self-identified as legally blind, their ability to perceive information visually varied. As such, we grouped them into two primary groups: group 1 included those who could not visually access the diagrams on the tablet, and group 2 included the rest who could perceive visual information from diagrams to a varying degree but largely relied on non-visual channels for information access.

Group 1 had 20 participants who ranged in age from 25 to 73, with a mean and median of 49.5, and SD of 13.59. 10 were female, 9 were male, and 1 was transgender. Group 2 had 5 participants with an age range from 35 to 101 and a mean of 61.6, a median of 58, and SD of 24.01. 4 were female, and 1 was male.

Participants' common ways of accessing diagrams included using screen readers for alternative text descriptions, asking others who are sighted, using tactile diagrams, and using visual aids like magnifiers. Educational diagrams and maps were the most common types of diagrams participants encountered. 21 participants mentioned that they did not often access diagrams or only access them if they needed to due to accessibility issues (N=7), or a lack of need or interest (N=5). 4 participants frequently accessed diagrams.

We recruited participants by advertising through various organizations such as the Canadian National Institute for the Blind (CNIB) and the Canadian Council of the Blind (CCB), and social media platforms. We advertised to previous participants from the interview study who consented to be contacted for future studies on diagram accessibility. We also recruited some via snowball sampling. In total, we recruited 25 participants with the following inclusion criteria: considering themselves to be totally or legally blind; relying on the auditory channel, or auditory combined with other sensory channels to access digital information; having experience with touch-screen devices; being 18 years and above of age. The detailed demographic information is summarized in Table 6.1. Participants were compensated CAD$30 for the one-hour in-person study session.

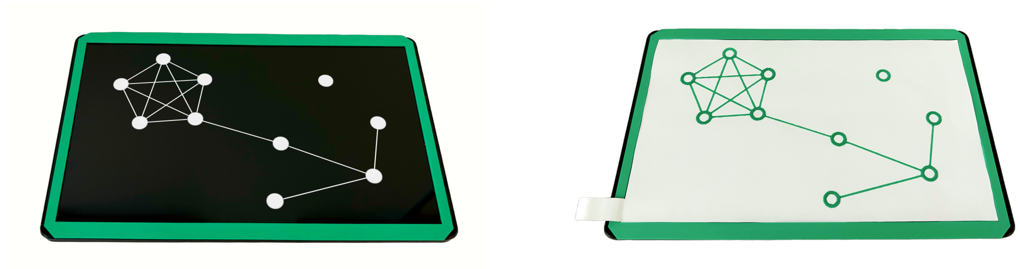
PID	Age	Gender	Educational Background	Professional Background	Degree of Vision Loss	Ability to Perceive Diagrams Visually	Onset of Vision Loss	Experience with Diagrams
P1	52	Female	High school	Massage therapist	Legally blind, less than 3% peripheral vision left	Not able	10 years old	Not often. Mostly educational diagrams in school.
P2	48	Male	Post-secondary	Construction business	Light perception in left eye, blurred vision in right eye	Not able	2022	Occasionally, especially transit maps.
P3	25	Male	Bachelor's degree in Computer Science and Business	Technology-related	Total	Not able	Birth	Daily encounter as a student.
P4	56	Female	Post-secondary	Senior management	Total	Not able	21 years old	Educational diagrams in school. Occasionally nowadays.
P5	46	Trans-gender	High school	Volunteer for non-profit organizations and local communities	Total	Not able	Birth	Occasionally, would like to avoid it whenever possible.
P6	53	Female	Bachelor's degree in English	Language teacher	Total	Not able	Birth	Every other day, quite frequently.
P7	70	Female	Diploma in Physiology	Program coordinator; physio-therapist	Total	Not able	5 years old	Frequently in education settings. Occasionally in professional work for reports containing diagrams.
P8	101	Female	Master's degree	Therapeutic dietitian	Legally blind, usable vision remaining in left eye.	Able to see diagram visually	Gradually losing, mostly in the last year	Occasionally
P9	73	Male	Post-secondary	IT technologist	Total	Not able	11 years old	Previously, educational diagrams in school, and flow charts in professional settings. Currently not at all.
P10	52	Female	Environment engineering	Manager	Legally blind, RP, no central vision, some peripheral vision left	Not able	18 years old	Previously almost all the time, work required to input data into reports, charts, diagrams. Currently, almost never because unable to.
P11	65	Male	High school. Currently taking college courses	Worker at a disability organization	Total	Not able	Born blind with light perception and colour perception. Totally blind when 7 years old	Tried to access diagrams in general, but never really worked. Not very often. Encountered maps in school.
P12	58	Female	Post-secondary	Unemployed	Legally blind, partial optic atrophy with peripheral vision	Able to visually perceive some	Prematurely	Only access if need to. Mostly mathematical diagrams in school.
P13	35	Female	Bachelor's degree in English	Program coordinator	Legally blind	Able to visually perceive some	Birth	Not often nowadays. Tactile diagrams of train stations. Educational diagrams in high school.

P14	41	Male	Technology-related	Technology-related	Light perception in left eye, total blind in right eye	Not able	Birth	Maps usually, if there is one available. Braille books. Tactile diagrams.
P15	56	Female	Master's degree in Music	Previously professional musician; After vision loss, teaching music and assistive technology	Legally blind, 3% vision in right eye.	Able to visually perceive some	18 years old. Gradually losing vision since birth	Regularly. Indoor maps.
P16	46	Female	Master's degree	Customer service and sales	Legally blind, totally blind in the right eye, some narrow vision in left eye	Not able	Birth	Sometimes. Diagrams from training materials. Educational diagrams in school. Not often nowadays.
P17	32	Female	Bachelor's degree	Government work	Total	Not able	Birth	Educational diagrams in school. Not often nowadays.
P18	50	Female	Diploma	Medical assistant	Legally blind	Not able	2013	Educational diagrams in school. Not often nowadays.
P19	58	Male	University	Accounting	RP, legally blind	Able to see diagram visually	2008	Rarely
P20	37	Female	Bachelor's degree	Non-profit sector	Legally blind	Not able	9 years old	Diagrams from professional development materials. Not very often generally.
P21	68	Male	Certificate from polytechnic institute	Previously transit driver	Legally blind, some peripheral vision left	Not able	2014	Do not know. Do not pay attention to them because of blindness.
P22	61	Male	Bachelor's degree in Engineering	Software engineering	Legally blind, some peripheral vision left	Able to visually perceive the diagram if closed up to the screen. Did not utilize vision to access diagrams during the study.	30 years old	Occasionally. Maps, general diagrams.
P23	38	Male	Bachelor's degree in Education and General Studies	Government employee; Salesperson; Worked in the sports and recreation industry	Total	Not able	Young adult	Quite often, few times a week.
P24	28	Male	Post-secondary	Assistive tech support	Legally blind	Not able	2016	Educational diagrams in school.
P25	49	Female	Bachelor's degree in Writing and Anthropology	Executive director at non-profit organization	Legally blind, can only see light and shadow	Not able	Birth, gradual vision loss	Hard to say because of not knowing about the existence of a diagram, or not telling if the graphics is an image of a diagram. Daily basis either at work or in personal settings.

Table 6.1: Details about participants from the system evaluation study.

6.1.2 Apparatus

We used a low-priced mid-range tablet (Lenovo Tab M10 Plus Third Generation with a 10.6-inch screen) running Android 12 and our TADA system for the study. For the study, we also connected the tablet to an external speaker to ensure the audio was clearly audible, especially at low pitch frequencies. A 3D-printed border was attached to the tablet screen to help prevent participants from going off of the tablet screen boundary, as seen in Figure 6.1a.



(a) Border to prevent touches (b) Tactile diagram for learning

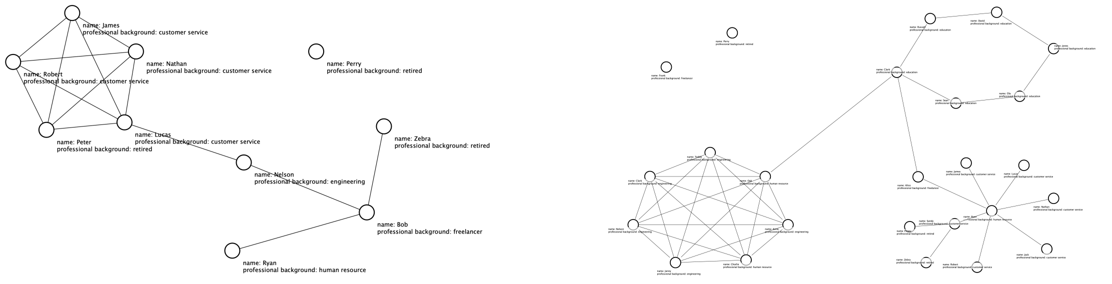
Figure 6.1: 3D-printed tablet overlays on the tablet used in the study

We also designed a 3D-printed tactile diagram overlay (Figure 6.1b), printed on a piece of high-quality printer paper cut out to be the exact size of the touch screen and could be attached to the tablet screen. This overlay was used to help participants to get familiar with our system and is a tactile representation of the digital diagram from Figure 6.2a. The overlay helped participants learn about a node link diagram by feeling 3D-printed rings that represented the nodes and lines that represented the link. The rings are raised slightly higher in height than the links to make them more distinguishable when touched. The tactile overlay was also used to teach participants the interaction techniques employed in TADA. This overlay works in conjunction with the TADA system running on the tablet, and the touch gestures are still recognizable with the overlay.

All study sessions were conducted in isolated private rooms. The experimenter was next to the table which participants sat in front of. We recorded the sessions and used a tripod with a camera pointing downwards to record participants' and interactions with the tablet.

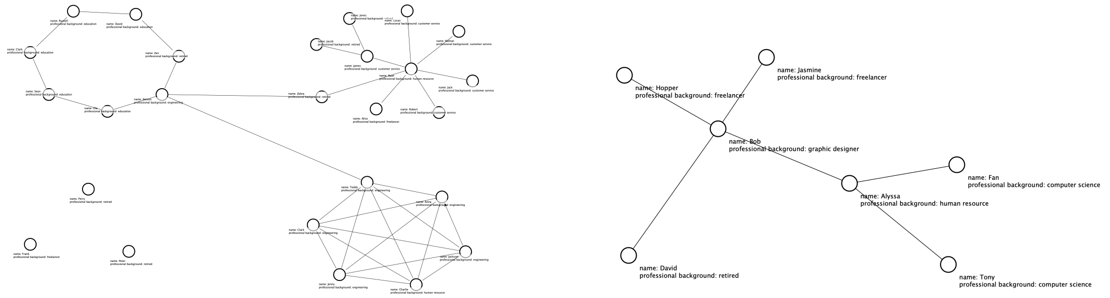
6.1.2.1 Stimuli

We designed four diagrams as the stimuli for participants to interact with in the study. They are friendship diagrams, where each node represents a person, and each link represents the existence of friendship between the two people connected to this link. We used this analogy of friendship as an easy way to introduce an example of diagrams to participants. Each person has some associated details or labels, such as their name and professional background. We divided two diagrams to consist of four quadrants: top left, top right, bottom left, and bottom right, to evaluate the effectiveness of using a single interaction technique to identify different characteristics of a node-link diagram in different spatial regions.



(a) Diagram for learning

(b) Test diagram 1



(c) Test diagram 2

(d) Test diagram 3

Figure 6.2: Diagram designs to be used as stimuli

The first diagram we created is the learning diagram, which was used to introduce participants to the system in every study session (Figure 6.2a). We designed this diagram to be diverse enough to showcase different kinds of connectivity and density of nodes across the diagram, yet not too complex potentially overwhelming the participants. In this learning diagram, there are people who are friends with each other showing high connectivity and forming a cluster on the left, with an empty space under this cluster. One person from the cluster has a friendship extending to the lower right region which is more sparse, with fewer people and friendships. There is one person on the upper right not connected to anyone.

Figure 6.2b shows test diagram 1, a more complex diagram than the learning diagram designed to evaluate the single-finger sweep interaction. The first quadrant (upper left) has few people with no friends. The second quadrant (upper right) has more people, and each is a friend with two other people in the same quadrant. The third quadrant (lower left) has even more people who are friends with each other, having the highest number of connections. The fourth quadrant (lower right) has the most number of people, with one popular person having high connections, and others who have only one or two friends. Quadrant 2 also connects to quadrants 3 and 4.

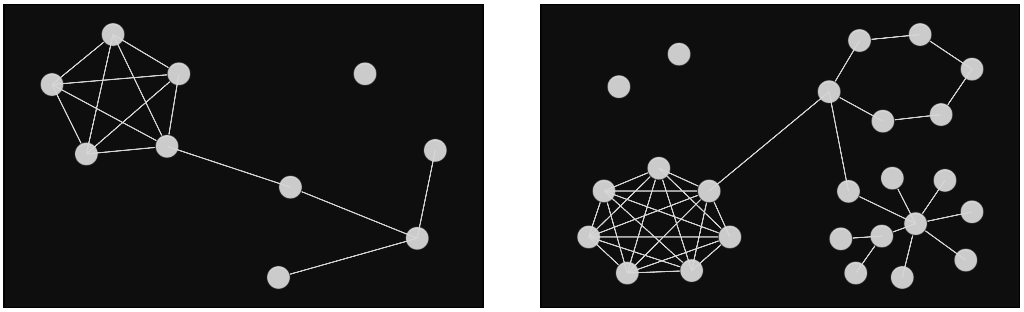
Test diagram 2 from Figure 6.2c has a similar complexity to the previous one to evaluate the five-finger dome interaction. To prevent participants from remembering answers from the questions asked for test diagram 1, test diagram 2 included modifications to test diagram 1 such as changing the positions and numbers of the nodes in each quadrant. Specifically, we added one more node to each of quadrant 1 and 2, removed one node from quadrant 3, and flipped the node positions in quadrant 4 in test diagram 1 while keeping the general spatial structure for the evaluation. We then shuffled the quadrants so that quadrant 1, 2, 3, and 4 from test diagram 1 now corresponds to quadrant 3, 1, 4, and 2 in test diagram 2. This way, we ensured that participants had to use the system to explore and understand each quadrant again, rather than relying on their previous answers.

Test diagram 3 from Figure 6.3d is designed to evaluate the remaining interaction techniques. We do not have to design different diagrams for the remaining interactions because the answers would not affect answering later questions.

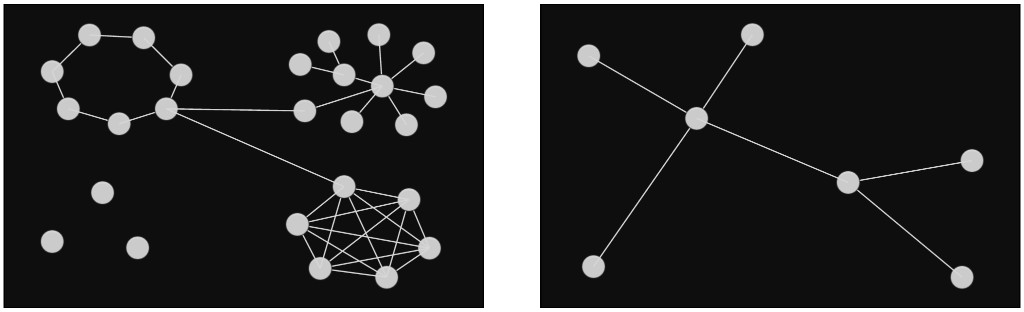
Figure 6.3 shows how these four diagrams are presented in the TADA system. The spatial information is preserved visually, and the other information can be retrieved using various interactions.

6.1.3 Study Procedure

At the beginning of each study session, we collected the participant's verbal consent for participation and demographic information including age, gender, background, vision loss, and experience with diagrams. We then provided a general context of the study, including the main goal of the system, what will be demonstrated, and a brief description of the learning diagram (Figure 6.1b, 6.2a, 6.3a). The main study was divided into 6 experimental tasks each focusing on an interaction technique. At the end of the study, we assessed participants' perceived workload using NASA-TLX. Lastly, we asked partic-



(a) Diagram for learning (b) Test diagram 1



(c) Test diagram 2 (d) Test diagram 3

Figure 6.3: Diagrams presented in TADA as stimuli

ipants about any feedback or comments they might have about the TADA system as a whole. The detailed study procedure can be found in Appendix B.

In general, the verbal consent and demographics data collection took about 5 minutes. Tasks 1 and 2 took 13 minutes on average. Tasks 3 to 6 took 6 minutes on average each to evaluate. The NASA-TLX questionnaire questions took about 5 minutes.

Before each task, we explained how the interaction and the corresponding audio work and guided participants through the learning utilizing the tactile diagram overlay described in Section 6.1.2. We also physically held their fingers to help them learn the gesture better only if they gave permission. We then ask participants to try the interaction themselves, first with the tactile diagram, then without it. We helped them with any difficulties they encountered and clarified any questions or confusion they might have. We then moved on to a different diagram for the task evaluation once the participant felt comfortable using the interaction.

The interaction techniques evaluated are (1) single-finger sweep; (2) five-finger dome; (3) dwell + tap; (4) single-finger flick-down, speech command for searching, and single-finger follow; (5) dwell + circle; and (6) dwell + radiate. We tasked the participants to only use the dedicated interaction technique to answer its associated question(s), and the questions asked about the test diagrams are listed below corresponding to the tasks.

We asked participants to think aloud while they were interacting with the diagram and considering the questions. At the end of each task, we asked the participants what they thought of the interaction technique.

Due to the constraint of time, we did not evaluate the interaction techniques designed for the features of filtering, listening to alt-text, and listening to the audio legend. The designed diagrams used in the study are not at the level of complexity requiring filtering to make themselves clearer. A general description of diagrams and what the audio means is informed by the researcher running the study to participants.

6.1.3.1 Task 1: Gain an Overview with Single-finger Sweep

We used test diagram 1 (Figure 6.3b) and tasked participants to gain an overview with single-finger sweep and answer the following questions about the test diagram:

Q1.1: Describe the overall layout of the diagram.

Q1.2: Which quadrant of the diagram has the most connections between people? (Answer: Bottom left.)

Q1.3: Which quadrant has the least connections? (Answer: Top left.)

Q1.4: Which quadrant of the diagram has the most people? (Answer: Bottom right.)

Q1.5: Which quadrant has the least people? (Answer: Top left.)

For the first question (Q1.1), we asked participants to sweep and explore the entire diagram, and provide answers regarding the general layout. We then asked more specific questions (Q1.1 to 1.4) to for participants to explicitly think in terms of quadrants and in terms of the connectivity and density of nodes. We asked participants to not count exactly but to gather a general overview.

6.1.3.2 Task 2: Gain a Summary of a Sub-region with Five-finger Dome

To learn about the effectiveness of the five-finger dome within sub-regions, we tasked participants to gather the summaries of sub-regions using five-finger dome and answer the following questions:

Q2.1: Which quadrant of the diagram has the most connections between people? (Answer: Bottom right.)

Q2.2: Which quadrant has the least connections? (Answer: Bottom left.)

Q2.3: Which quadrant of the diagram has the most people? (Answer: Top right.)

Q2.4: Which quadrant has the least people? (Answer: Bottom left.)

Q2.5: Which quadrant has people with unequal numbers of friends? (Answer: Top right.)

We asked the same specific questions about connectivity and density of nodes, but switched the diagram to test diagram 2 (Figure 6.3c) to make sure participants cannot reuse the same answers from test diagram 1. Additionally, because we also want to evaluate explicitly the effectiveness of the pacing of audio, Q2.5 was asked to see if participants could pick up the variations in the pacing of the link audio sounds within an area. There can be multiple answers to Q2.5, and we asked participants to identify the most obvious one where there is one person very popular in the top right quadrant while the others are not so that this node's link sounds are paced much faster. Again, we asked participants to not count exactly but to gather a general overview, and we did not evaluate what details this interaction could potentially achieve.

6.1.3.3 Task 3: Gain Textual Details with Dwell + Tap

We asked a question about the details of a diagram element from performing the dwell + tap interaction:

Q3: Given a person, what is the name and professional background associated with it?

We used test diagram 3 (Figure 6.3d) for this task and the remaining three tasks (Task 4 to Task 6). In Q3, we provided participants with a node to examine that had a name other than "Bob" who was going to be used in later questions.

6.1.3.4 Task 4: Search with Single-finger Flick-down, Speech Command, and Single-finger Follow

We tasked the participants to locate a piece of information and answer the following question:

Q4: Where is Bob? Locate Bob.

To eliminate other factors affecting the accuracy of speech recognition such as participants' accents and background noise, we conducted a Wizard of Oz experiment where we asked participants to only look for where Bob is in the diagram, and the system is only configured to search for Bob.

6.1.3.5 Task 5: Explore Connected Links with Dwell + Circle

We asked participants to perform dwell + circle to identify the number of connected links of a node:

Q5: How many friends does Bob have? (Answer: 4.)

Dwell + circle aims to explore a node's connected links in detail, gathering the number of connections, their directions, and their relative distances. Q5 only asked about the number of connections because, in the interest of time, we decided to focus on if one could complete a full circle and count the number of connected links without exploring other details. Q6 below partially explored the directional information of the connected links.

6.1.3.6 Task 6: Navigate with Dwell + Radiate

Following Task 5, we asked participants to navigate to any nodes of the neighbouring node using dwell + radiate:

Q6: Navigate from Bob to any of his friends.

Dwell + radiate works in combination with dwell + circle. Using the dwell + circle interaction to explore the connected links, one can locate a link of interest, and then decide to navigate towards its direction using the dwell + radiate interaction. Therefore, Q6 evaluated the dwell + circle interaction partially, and the radiate interaction.

6.1.4 Data Analysis

We employed a mixed-methods approach, using both quantitative and qualitative data sources. We collected data from the task outcomes and answers to task questions, the NASA-TLX questionnaire, the participants' think-aloud comments, and their feedback on the system. We analyzed the data using descriptive statistics, box plots, and affinity diagramming [55] to answer the research questions.

As described previously, we divided our 25 participants into two groups: Group 1 consisted of participants who were totally blind, or legally blind and could not visually access diagrams (N=20) and Group 2 included the rest who was legally blind and could visually access diagrams to varying degrees (N=5). We analyzed the data separately for the two groups of participants.

Most of the questions asked were quantitative and assigned the binary answers of "Correct" or "Incorrect". During the study sessions, some participants needed some assistance performing what they were asked to do. Such scenarios were noted to be "Correct with assistance". These data were compiled into statistical summaries showing the number of participants for the degrees of correctness for each question. The NASA-TLX scores were also summarized into box plots.

For questions that require qualitative answers such as Q1.1, and questions related to what participants thought of the interaction techniques and the TADA system, these answers were transcribed from the study sessions and grouped into common categories with the help of affinity diagramming.

6.2 Results

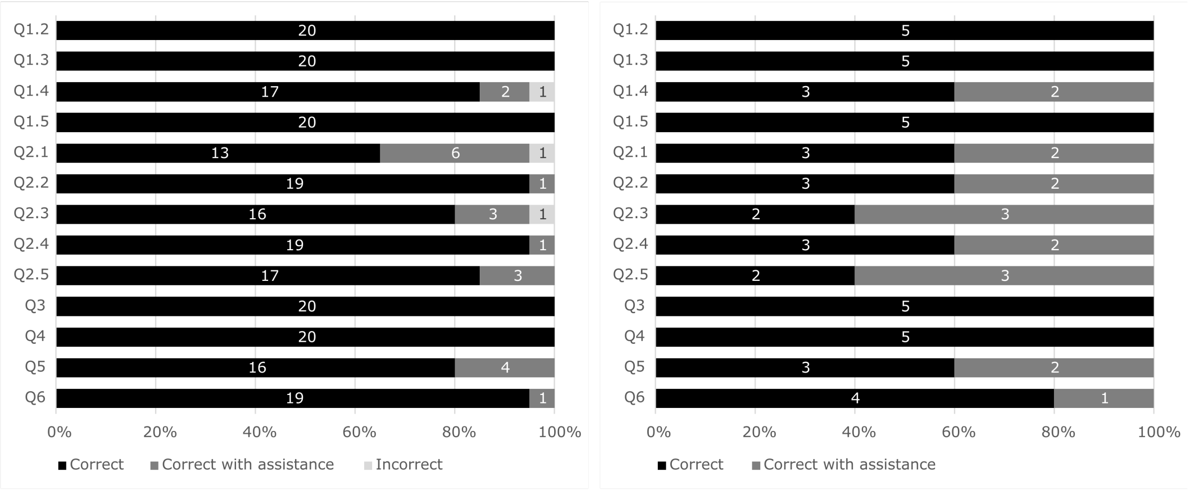
In this section, we present the task and question outcomes and perceived workloads in terms of the two groups of participants, and the more qualitative findings under the subsection of participant feedback.

6.2.1 Task and Question Outcomes

Q1.1 required qualitative answers and we grouped participants' responses into four levels of detail. The basic level is where participants could tell that no regions are empty in the diagram, and that there are nodes throughout the diagram. Both groups 1 and 2 went beyond this basic level.

Level 2 includes some specific details such as the connectivity or density of nodes. Example quotes from participants include *"... on the left ... there's a lot of connections over here ... That one has ... [no] lines with it [on the upper left] ... Here, it's more sparse [on the right] ..."* [P1 from group 1], and *"That's a group of people [on the lower left] ... [There're] single individuals [on the upper left]."* [P8 from group 2].

Level 3 includes specific details such as both the connectivity and density of nodes, and this level is where we expected most of the participants would be able to achieve at.



(a) Group 1 quantitative results

(b) Group 2 quantitative results

Figure 6.4: Correctness of answers to quantitative questions listed in Section 6.1.3

Example quotes include *"What I'm sensing here is the bottom left has more connections. There are connections on the top right and [bottom right] but they are more widely spaced apart ... At the top left, there are two [people] but they are not connected. They don't feel like they are as connected."* [P16 from group 1], and *"There's quite a few people over here [on the lower left] ... [On the lower right, there's] a bunch of people and guitars ... There's a lot of connections but there is not that many people. It's more people over here [on the lower left] ... and over here [on the upper left] there are two people ... They are not really that far apart but they're not connected."* [P12 from group 2].

The top level contains extra details such as connections between quadrants, and characteristics of distributions or connectivity of nodes. Examples of such extra details include *"[On the lower right,] there is a person in the middle, I think, that's connected to all other people around ... as I am going in circle here."* [P6 from group 1], and *"[On the upper right, each person is] connected to two people [in a way that they] are not connected with people across the circle."* [P15 from group 2].

For the remaining 13 questions, their quantitative scores are summarized in Figure 6.4. For both groups, Q1.2, Q1.3, Q1.5, Q3, and Q4 had all correct answers without any assistance. For the remainder of the questions, participants asked for some clarifications or assistance from us and these can be categorized into five types. First, we clarified the difference between the density of nodes and the density of links because some participants thought high connectivity mean a high number of nodes which does not hold true all the time (for example, a group of isolated nodes do not have any connections), in Q1.4 and

Q2.3. Second, we assisted with explaining the question further because some participants were unsure about the meaning of the term “people with unequal numbers of friends” from Q2.5. Third, some participants needed some help correcting their gesture(s). Fourth, some needed guidance on covering a quadrant more completely with the five-finger dome interaction. Lastly, there were instances when participants gave an answer to Q2.5 that could be mistaken for the correct one. In such cases, the researchers reminded the participants about what the question was specifically asking for to help them provide a more accurate response.

6.2.1.1 Group 1 Results

For Q1.1, a majority of the participants were successful in providing specific information about the connectivity and density of the nodes. Group 1 (N=20) had 0 participants at the basic level, 2 participants at level 2 (10%), 15 participants at level 3 (75%), and 3 participants at the top level (15%).

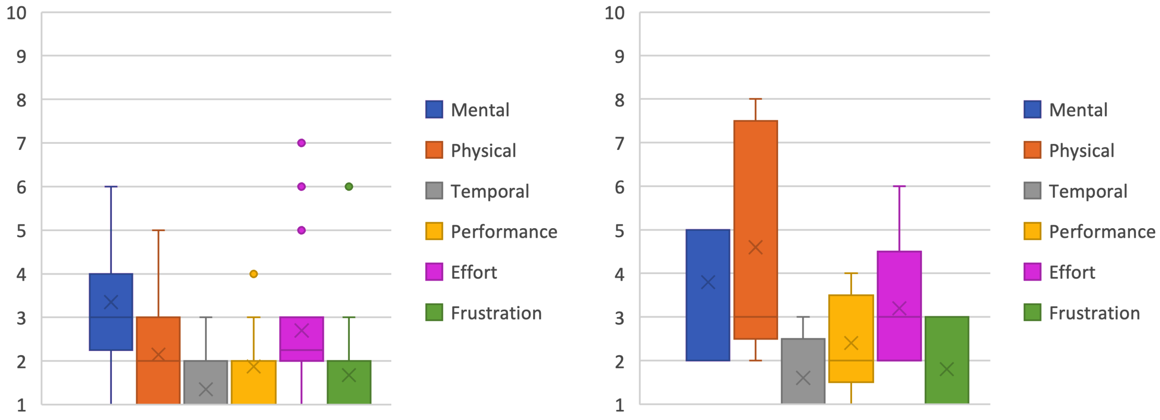
For the remaining questions (Figure 6.4a), participants provided mostly correct answers with only 3 incorrect ones in Q1.4, Q2.1, and Q2.3. Some correct answers needed some assistance. Here we explain why some participants needed assistance or did not get correct answers. Q1.4 and Q2.3 are the same questions on identifying the quadrant with the most number of nodes. 3 participants in Q1.4 and 2 participants in Q2.3 believed that the quadrant having the most connections is the one with the most nodes. They needed clarification on the difference between the density of nodes and the density of links. 2 participants in Q1.4 and 1 in Q2.3 got the right answer with this clarification. 1 participant in each of Q1.4 and Q2.3 did not get the correct answer in the end.

In Q2.1, participants were asked to use the five-finger dome interaction to identify the quadrant with the most connections. Five participants were confused about the quadrant where one node is very connected to others which are not as connected. For example, in Figure 6.3c, the top right quadrant had a popular node which generates fast-paced link audio only for this node. Some thought this meant the answer, and we reminded them to consider all connections in this quadrant and their link audio, not just from one node. With this reminder, 4 participants answered correctly and 1 did not. In Q2.5, the question itself confused 2 participants who needed more explanation on the meaning of nodes having different numbers of connections and they got the correct answer afterwards. For Q2.1 to Q2.5 meant to evaluate the five-finger dome interaction, 2 participants needed assistance in covering a quadrant more completely using the dome gesture. 1 of them needed help with all 5 questions, and the other participant needed it occasionally in Q2.1 and Q2.3. In Task 5 which evaluated the dwell + circle interaction, 4 participants needed assistance with performing the gesture due to the dwelling finger drifting. In Task 6 which evaluated the dwell + radiate interaction, 1 participant moved their radiating finger too quickly and lost track of the direction. The participant was able to acquire the correct answer after helping their gesture move more slowly.

6.2.1.2 Group 2 Results

For Q1.1, group 2 had 1 participant at level 2, 1 participant at level 3, and 3 participants at the top level. For the quantitative questions (Figure 6.4b), all participants got correct answers, and some needed assistance to acquire the right answers in terms of clarification of concepts and gesture correction. Particularly, everyone got Q1.2, 1.3, 3, and 4 correct without any assistance. For Q1.4, 2 participants needed clarification of the difference between the density of nodes and density of links. In Task 2, the five-finger dome gesture needed assistance in terms of helping the fingers stay on the screen consistently while keeping the hand to be in a dome shape. 2 participants needed help for all questions in Task 2 due to difficult coordination of fingers, and 1 participant only needed help with gesture in Q2.3. For Q2.5, one participant needed help with understanding the question. In Task 5, 2 participants asked for help in supporting their gesture, and one of them also needed it in Task 6.

6.2.2 Perceived Workloads



(a) Group 1 NASA-TLX results

(b) Group 2 NASA-TLX results

Figure 6.5: Box plots of NASA-TLX perceived workload ratings.

We collected perceived workload ratings for the TADA system using the NASA TLX questionnaire which asks participants to rate the workload of tasks on six scales: mental demand, physical demand, temporal demand, performance, effort, and frustration [41, 40]. The response scale typically has 21 marks on a line, but because we were verbally asking participants about the questions, we modified the scale to be from 1 to 10. Lower scores correspond to less perceived workload. Group 1 (Figure 6.5a) had mean scores of 3.35, 2.15, 1.35, 1.88, 2.7, and 1.68 for the six dimensions respectively. Standard deviations were 1.35, 1.31, 0.59, 0.81, 1.59, and 1.26. For group 2 (Figure 6.5b, the mean scores were 3.8, 4.6, 1.6, 2.4, 3.2, and 1.8. Standard deviations were 1.64, 2.70, 0.89, 1.14, 1.64, and 1.1.

In general, we see higher mean and standard deviation scores for mental demand, physical demand, and effort for both groups. Some participants who gave higher scores in these categories commented that these could be because of the mental efforts needed to understand and interpret the audio information (N=4) and rarely the questions asked (N=1), and the physical efforts from conducting certain gestures (N=2). The effort score is generally a combination of mental and physical demands and correlates to them. The higher standard deviations show that some participants found the system to be more effortless to use than others. Group 2 generally had higher scores than Group 1 possibly due to its small sample size minimizing its statistical influence. P8 from group 2 mentioned that they needed to work hard to understand the questions due to an old age of 101. In group 2, the physical demand scored higher likely because it had more instances where people needed assistance with performing the gestures (14 total instances versus 5 in group 1). In some cases, this was due to a lack of experience with performing gestures (P8), and personally low dexterity of fingers (P12).

6.2.3 Participant Feedback

From participants' thinking aloud, answers to the questions on their thoughts on the TADA system, and observations of them using the system, we incorporated their qualitative feedback in terms of the effectiveness of the interaction techniques and the audio feedback. Overall, P2, 11, and 16 called TADA to be *"fun"*; P3, 4, 5, and 22 called it *"intuitive"*; P1, 3, 4, 5, 7, and 11 called it *"useful"* or *"helpful"*; P1, 18, 22, and 25 called it *"innovative"*.

6.2.3.1 Audio Information

Below we report on participants' feedback related to perceiving auditory information.

Distinct Usages of Different Non-speech Audio and Speech: Multiple participants noted explicitly (N=8) that the sounds meant for different pieces of information were distinct to be made sense of. P1 and P25 commented that the audio was of good quality. P10 commented that the non-speech sounds and speech were effectively utilized for different purposes: *"When you are just gathering a sense of the general layout, I think the sounds are perfect because if [the system] starts to tell you [such information in speech] all at once, you cannot distinguish [the information]"*. Speech is used for describing more details and instructing directions for searching which *"were straightforward"* (P10).

More particularly, participants commented that the audio from the single-finger sweep (Task 1) was informative to gain an overview of a diagram (N=16). The different pacing of link sounds using the five-finger dome (Task 2) informed about the variety in the nodes' connections within a sub-region in a diagram (N=11). 4 participants explicitly noted the appropriate usage of speech to gather detailed information using the dwell + tap interaction (Task 3).

Positive Feedback on Searching Speech Feedback: In Task 4, P9 and P21 commented that the pacing changes in the directional speech instructions from the single-finger follow was *"logical"* and *"helpful"*, informing about the distance from the finger to the target location. P23 also found the directional responses in speech to the finger movement to be efficient, allowing P23 to identify the target quickly.

Positive Feedback on Navigation Audio: In Task 5, the proximity guidance audio with varying volumes used for the dwell + circle interaction helped locate the connected links of a node (N=4), and P10 could also use it to infer the presence of a link although the finger did not cross the link. For the dwell + radiate interaction (Task 6), the constant pluck-like audio from the radiating gesture interacting with the link helped participants stay on track (N=7). Additionally, 5 participants also noted that the proximity guidance audio also helped them stay on track, and 6 participants noted it informed them about the proximity of how close the target node was to their finger position. P18 commented that the speech associated with Task 6, "Navigating", served as a clear indication of moving from dwell + circle to dwell + radiate, similar to the blister-like bumps on sidewalks in real-life to discern where a sidewalk changes to a crosswalk for navigation.

Negative Feedback on Audio Summary from Five-finger Dome: However, some participants (P2, 10, 11) found the audio summary from the five-finger dome interaction (Task 2) to be too much information making it hard to process. In Q2.1, five participants confused the fast pace of connected links of one node with the whole quadrant having the most connections. Four of them acquired the correct answer after some assistance in reminding them that the question asked about the whole quadrant, and one still did not get the correct answer.

6.2.3.2 Interaction Techniques

Below we report on participants' feedback related to the use of interaction techniques.

Difficulties with Five-finger Dome and Dwell + Circle Interactions: From the quantitative results, we know that some participants needed assistance conducting especially the five-finger dome (Task 2) and the dwell + circle (Task 5) interactions. P8, P13, and P16 commented that the five-finger (Task 2) gesture felt too clustered and awkward to perform. P16 also had long nails which made the gesture hard to perform as they get in the way of touch recognition, and this was mentioned by P6 and P15 as well.

For the dwell + circle interaction (Task 5), 6 participants found it to be a bit awkward. The main issue was the dwelling finger drifting as participants performed the circling gesture using another finger. P25 commented that this was due to the dwelling finger pressing too hard onto the screen. Because the node sound gets muted during this interaction to prioritize the proximity guidance audio, participants might not notice the drifting unless they lift up and down either the dwelling finger or the circling finger to re-trigger the node sound to confirm they are still on the node. Interestingly, the finger drifting issue did not impose as a problem to the same extent in other interaction techniques involving dwelling (no failure or assistance needed on this). From what we observed from the participants' interactions, the drifting interfered with the relative positioning of the two fingers in the dwell + circle gesture, resulting in a less accurate interpretation of the directions of links coming from the node being dwelt on.

Effects of Tablet Familiarity on Interaction Performance: Such difficulties in interactions could also be affected by the fact that some participants were less used to the tablet. P2, P4, and P25 commented that personally owning the tablet would help them better position their fingers for more accurate interactions.

Positive Feedback on Other Interactions: However, there were also others who commented that these interactions were easy to conduct (P9 and P19 for the five-finger dome (Task 2); P14, P23, and P24 for the dwell + circle (Task 5). P3, 5, and 18 also found the dwell + circle interaction to be informative of the directions of connected links. For other interactions, at least seven participants found the dwell + tap interaction (Task 3) to be easy. P6 also commented that repeated tapping while dwelling was *"stable"* and *"consistent"*.

For Task 4, participants liked the use of speech recognition. P22 mentioned that this prevented the need to type on a touch-screen device which could be troublesome, and P18 commented that this provides more ways for access to information. For the dwell + radiate interaction (Task 6), P3, 16, and 23 said it was straightforward and easy to use.

Effects of Individual Differences: Generally, P12 commented that the success of answering the questions would depend on participants' individual skills and abilities. For example, P2, 11, and 12 said they had less superior motor skills affecting their performance with physical gestures.

Chapter 7

Discussion

In this chapter, we discuss the evaluation results, and how TADA meets the design goals. We also delve into its design approaches and trade-offs, compare it with other accessible diagram formats, assess its limitations, and propose future work. The DPs are discussed throughout the chapter because we want to show how they are applied in different aspects of TADA's design. Participants referred to in this section are from the system evaluation study.

7.1 Discussion of Evaluation Results

The results from our evaluation study in Chapter 6 showed that most participants were able to successfully perform the tasks and answer the questions correctly from the audio feedback using the interaction techniques supported by TADA, with some assistance for certain clarifications or gestures. To more exactly answer RQ1 from the evaluation study, participants found the single-finger sweep, dwell + tap, and single-finger follow with speech commands to be easy and intuitive. The five-finger dome, dwell + circle, and dwell + radiate sometimes needed more practice and precision. The audio feedback was generally clear and informative with some participants complimenting the distinctness of audio used for different purposes (DP1), and the audio quality. The audio from the single-finger sweep was effective in providing an overview of the diagram, while the five-finger dome was useful in summarizing a sub-region as another way of gaining an overview; The dwell + tap helped participants access textual details of a diagram element; The speech interaction and single-finger follow were straightforward and convenient for searching; The dwell + circle was informative of the directions and distances of connected nodes; The dwell + radiate was straightforward and easy to navigate.

These findings for RQ1 suggest that TADA supports accessible and feasible interaction techniques with audio feedback for non-visual access to node-link diagrams. We found that participants could use different interactions to access different levels and perspectives of information (DP3), from overview to detail, from global to local, and from spatial to semantic. Participants could use the different modalities supported to interact with the system including touch gestures and speech commands. We also learned that participants could benefit from the audio feedback that conveyed various aspects of diagram information, such as connectivity, density, direction, distance, location, and textual details.

For RQ2, participants found the mental demand, physical demand, and effort needed to be the highest factors of their workload perception. This could be due to the cognitive and motor skills required to understand and interpret audio information, and perform gestures accurately. While we did not use the NASA-TLX questionnaire to understand the exact differences between the experimental tasks, it appears that higher mental demand scores are driven by tasks or questions that involve more complex reasoning or decision-making compared to those that involve more straightforward or concrete judgments. This was particularly true for the five-finger dome interaction. Similarly, higher physical demand scores are driven by interaction techniques that require more precise or coordinated gestures such as the five-finger dome, and dwell + circle. We will discuss and address these system limitations later in the chapter.

7.1.1 Limitations of the Evaluation Study and Future Studies

Our evaluation study focused on assessing the key interaction techniques of TADA that are essential to its operation. However, due to the nature of such a study design, we were unable to comprehensively evaluate all features, including some less central but still important ones such as the filtering mode, alternative text support, and audio legend.

Additionally, we did not evaluate the use of multiple interaction techniques simultaneously or compare different techniques for performing the same task or accessing the same information. The evaluation study so far is only preliminary because we think it is currently more important to evaluate if the basic tasks are achievable with the supported interactions. Because the evaluation focused on the use of interaction techniques in isolation, we miss the bigger picture of whether participants would be able to effectively combine and switch from one to the other (therefore not evaluating DP5 explicitly). This definitely needs further study, and we take a bottom-up approach to conduct higher-level evaluations in the future. We could also test our system with more complex or diverse diagrams and visualizations to evaluate its performance and identify potential design challenges or opportunities.

7.2 Meeting the Design Goals

This section discusses the extent of TADA meeting the two design goals from Section 4.1.

7.2.1 DG1: Achieving Higher Levels on the Ladder of Diagram Accessibility

In Section 3.3.1, we conceptualized diagram accessibility as different levels on a ladder, from the lowest to the highest. It starts from not knowing the existence of a diagram, to knowing a placeholder of a diagram, to accessing a static single view of a diagram, to accessing multiple perspectives of a diagram, to conducting dynamic queries about a diagram, and finally to achieving complete and equal access to diagram data. Here, we use this same ladder to understand in which ways and to what extent TADA provides access to diagrams. We discuss this from the top level going down to highlight the more important contributions.

At the highest level (six) of access, TADA aims to present complete diagram information, within the current scope of a node-link representation of diagram data. We try to present the spatial and semantic information effectively by designing various appropriate interaction techniques and audio feedback to support people with gathering overviews and details, and conducting operations such as searching, navigation, and filtering. Additionally, TADA provides people with a larger degree of agency to make choices and decisions on what information to access and what to do with the information than some other ways of access. For example, one does not need to be dependent on others' subjectivity as part of their description of a diagram, or the extent of information statically presented by textual descriptions. Multiple ways of interactions are provided to access multiple levels and perspectives of information as options (DP3) for people to choose to process the information and conduct various possible tasks. P13 and P25 commented that the system gives them a higher level of confidence and independence. P20 considered this to be a helpful tool in the toolbox of accessibility as another option to utilize depending on the situation, and P25 commented that TADA provides ways of access which current technology does not support. The level of understanding of diagram information however still depends on the individual differences in people's ability and experience; TADA supports people to achieve equal access, but the actual outcome of people's understanding would vary.

TADA provides dynamic queries by supporting interactions and audio feedback, meeting level five on the ladder of diagram accessibility. People query TADA to access specific information from a diagram by conducting gestures or speech commands; TADA responds in real-time audio feedback which guides people towards further engagement with their interactions, creating a continuous and interactive experience. Moreover, by providing people with a larger degree of agency, people could acquire a deeper level of knowledge and insight. This can be achieved by using possible combinations of interactions and their corresponding audio feedback. These are not evaluated in our study, but some participants explored such combinations themselves (DP5). For example, P20 commented that while using the five-finger dome interaction to learn about a sub-area in a diagram, one could combine it with the single-finger sweep within the same area to gain more informed knowledge. P13 discovered that one could put down multiple fingers in different parts of a diagram to compare and contrast. P13 put them on two different nodes and lifted one of them to hear the other clearly, then switched back and forth to identify the difference in pitch to know which node is more connected. Similarly, P24 could use the dwell + tap interaction within the local areas of two nodes in parallel to listen to their individual textual details. Multiple participants were able to naturally utilize the dwell + tap interaction once learned and combined it with other interaction techniques in the subsequent experimental tasks to gather more information about the nodes they were interacting with. By utilizing these different interactions and combining them in various ways, people can gather more information and gain a deeper understanding of the diagram from multiple perspectives.

TADA presents multiple perspectives of a diagram depending on the use of interaction techniques, meeting the fourth level and going beyond the third level on a single static diagram view on the ladder. Firstly, one single interaction technique supports multiple tasks (DP3). For example, the default mode of TADA is using single-finger sweeps to explore a diagram. Quicker sweeps over nodes and links generate more discrete and faster-paced sounds constructing an audio texture as an overview, and slower sweeps help people explore the individual details more clearly. With the five-finger dome, people can adjust the size of the dome for a summary of more information by covering more area, or focus on a slower audio summary to listen to more details by closing in their fingers.

Secondly, multiple interactions can help access the same type of information (DP4). For example, the single-finger follow is designed to find the spatial location(s) of node(s) of interest by explicitly telling the system about what to search. The single-finger sweep also inherently searches for spatial locations of nodes or links of contrast like a highly connected node, and dwell + circle also informs the spatial directions of connected nodes. For filtering, the five-finger dome, single-finger follow, dwell + circle, dwell + radiate, and the exclusive filtering mode all support filtering in different ways by leaving out nodes and links that are not necessary for the tasks.

The bottom two levels on the ladder of diagram accessibility are out of the scope of the current design of TADA. We assume that there already exists diagram data in a semantic and machine-readable format, which is parsed to TADA to generate the spatial representation either from the data or using the force-directed layout at level three on the ladder. There exists prior work that achieves the automatic extraction of diagram data from visual diagrams [57, 60, 76], and this is part of the future work for TADA.

7.2.2 DG2: Reducing Barriers to System Access

We chose to use a tablet due to its relatively low cost and general wide availability as compared to tactile formats which could be expensive and time-consuming to produce, and we aimed to use a low-priced mid-range tablet to further drive down the economic cost. Anecdotally, we received multiple comments that suggest having a low cost is the right direction of design from the formative study (See Section 3.2.2.2 and 3.2.2.3 related to the drawbacks of other forms of access and tools).

We retained standard touch interactions such as holding or dwelling, swiping, and tapping to keep the learning curve minimal. P19 found the gestures to be familiar from common touchscreen devices. P24 also pointed out that the circling gesture in particular is similar to the "VoiceOver rotor" [8] supported on iOS devices. Our interaction techniques are combinations of these with new additions, but we aligned with existing work as much as we could. We borrowed the flicking gesture and second-finger tap selection from the work by Kane et al. [53], but instead of tapping anywhere on the screen, we restricted the tapping to be within the local area of the dwelling finger to support the simultaneous exploration of multiple diagram elements with separate hands. For interactions involving at least two fingers, people can also choose to use only one hand or both hands to conduct them. Some participants found using one hand is easier for certain interactions, and some preferred using two hands. We made the dwelling to be a pre-existing condition for subsequent interactions from a second finger because we want people to maintain their spatial position (DP2) and because it is common to keep a reference point as people read braille or tactile maps, as informed by literature [31, 89] and our previous formative study. The threshold values used to detect and recognize touch gestures were tweaked and refined by trials and errors during the implementation phase, and by conducting informal pilot studies with three participants to ensure the detection is optimal.

We learnt from the study by Ramôa et al. [80] that voice-based guidance is the fastest for people to pinpoint a location and does not require training. We considered this finding while implementing the search mode. In addition, we also varied the pacing of the speech to help inform about the distance towards the target, and introduced two ways of finger movement: one can choose to either move diagonally, axially, or mix both and the audio feedback adjusts dynamically. P6 and P23 from the evaluation study liked the freedom to choose between the options. 10 participants commented that the use of voice-based guidance was very *"straightforward"*, and they did not need to memorize any information. P24 found the system was *"sort of thinking for me"*. We also utilized the native speech engine from the mobile platforms so that people are already familiar with it and existing customization to the synthesis engine could be applied as part of future work. P19 particularly liked the familiarity of the speech engine used in a typical screen reader, and P22 also mentioned that the usage of speech in TADA aligns with the habits of screen reader users.

We used real-life metaphors in some of the designed audio feedback. The guitar sounds were meant to represent the links, and if a link is crossed, a plucked guitar sound plays as if a guitar string is plucked in real life. Moreover, the proximity guidance audio is similar to submarine sound navigation which is also used in the sonar navigation in the work by Ramôa et al. [80] but without the usage of the pitch which is already mapped to other information in TADA. Some other aspects of our audio design were also observed by participants to resemble real-life scenarios. For example, P2 found the audio texture from the single-finger sweep to remind P2 of the physical texture of a street surface while using a white cane. P18 also mentioned the single-finger follow interaction during searching is similar to using a cane to navigate in real life following directions. P18 mentioned that the "Navigating" speech feedback from the system when conducting dwell + radiate reminded P18 of the sidewalk bumps in real life where a sidewalk merges to a crosswalk.

Although TADA focuses on non-visual access to diagrams, it still represents diagrams visually. Some participants who were legally blind and could visually access diagrams to different degrees commented on the visual access: P13 who was able to visually perceive a diagram partially mentioned that the contrast of the white diagram with a black background was beneficial; P13, P19, and P22 all mentioned that people with low vision are likely to utilize both visual and audio information to access diagrams, adjusting to their best abilities.

7.3 Design Approaches and Trade-offs

In this section, we discuss some of the design approaches and trade-offs that we considered when creating TADA.

7.3.1 Benefits and Challenges Associated with Preserving Spatial Information

With TADA, the interaction techniques are designed in such a way that people can inherently understand the spatial information of a diagram while interacting with it (DP2), either by touching diagram elements spatially or following the system's directional instructions. Multiple participants were able to form a mental image of a diagram or a sub-region of it, using a common strategy of holding the tablet with one hand to better relatively position the other hand for interactions. Both the single-finger sweep and the five-finger dome techniques effectively conveyed overview information from the spatial area being interacted with as most participants succeeded in Tasks 1 and 2. P1, 3, 14, and 24 particularly noted the usefulness of the five-finger dome interaction to provide a high-level understanding of the information within a defined spatial area. P6 commented that the audio feedback is repeatable and stable which helped with building a picture in the mind. For dwell + circle, P18 commented that the pitch of the link audio informed how far the connected nodes were, and P16 and P18 noted the proximity guidance audio was effective to map out the empty space between the links, helping them form a mental image of the space interacted. With the single-finger follow interaction, P18 and P25 said that they could visualize the path of movement toward the target. P18 particularly commented that even though the system used a flat touch screen, it has the *"tactile nature"* because the audio managed to inform the necessary spatial information to help people form a mental visualization of the diagram.

For interaction techniques that involve dwelling, P12 and P18 commented that they liked keeping a finger on a spot and then gathering more information locally because this helped them to not lose their position, as per DP2, suggesting this design was effective. P6 also mentioned that this felt familiar to reading braille which involved keeping a finger as a reference point. This also aligns with one of the guidelines for non-visual graphics access [89], which suggests the use of multiple fingers as reference points.

However, some participants found the dwelling to be awkward especially when combined with another finger circling around the dwelling finger, and some encountered the issue of finger drifting without being aware of it happening. This is not a direct drawback of DP2, but one of the interactions related to spatial information. We expand the interactive area of a node as a finger dwells on it to help resolve the issue to some extent. Some suggested borrowing techniques from screen readers on mobile like Voiceover on iOS [9] and Talkback on Android [7], which provide a focus box on an element on the screen instead of using a finger to hold it down. This aligns with one of the guidelines from McGookin et al. [71] that recommended avoiding interacting on specific spatial locations. While this design reduces the physical demand, people would have to mentally remember the spatial position of a screen element. Icons and menu items typically have a grid-like or list-like structure to them making it easier to remember, but when it comes to diagrams which do not have a standardized structure, remembering where the nodes and links are can be challenging. Dwelling on a diagram element to interact with and explore it, therefore, has its benefits. As part of the future work for TADA, we could introduce another interaction (like dwell + double tap) as an option to allow people to lock and unlock the focus onto a particular diagram element for people who do not want to dwell + other continuous interactions in some cases.

7.3.2 Balancing between Information Access and Information Overload

We aim to help people move up the ladder of diagram accessibility by providing multiple perspectives of information in a dynamic way (DP3), but sometimes the information might be too much to interpret especially if a diagram gets too complex. For example, P2, 10, and 11 found the audio summaries using the five-finger dome to contain too much information for them to process. However, this is the only interaction technique in which participants found it difficult to decipher the audio. Participants were able to get most of the correct answers to the questions from the tasks.

There are also strategies such as providing an option to slow down the audio summary suggested by P7 and P23, or closing in the five fingers to hear less information from a smaller area. Moreover, DP4 helped provide flexibility and redundancy for accessing the same type of information for people to choose and combine interaction techniques. P25 commented that TADA was able to portray multi-dimensional information using audio and interactions in a non-linear way, and P25 appreciated that TADA provided a new way to perceive and learn complex information using audio. Part of the reason why some participants found the audio hard to process was also because no extensive training was provided during the evaluation study. P7 mentioned that learning everything in one hour was hard, and most participants commented that perceiving the audio information would be easier if they had more time to practice and get more familiar with the system.

7.3.3 Balancing New Concepts with Prior Knowledge

We aimed to minimize the learning curve as part of DG2 and this is largely discussed in Section 7.2.2 on the discussion of how TADA meets DG2. Additionally, we strived to keep the basis of the meanings of audio information consistent for the different interactions as foundational knowledge to help people learn new concepts (DP1). For example, the node and link audio from the single-singer sweep forms the audio summary from the five-finger dome. For dwell + radiate, we did not design different audio for a finger to stay on track during navigation but utilize a wiggling motion to continuously move across the link to produce discrete plucky sounds to inform people they are on track. The proximity guidance audio always informs the distance toward the target(s) of interest. Speech interaction is used for triggering particular modes and is distinct from the use of non-speech audio.

In retrospect, we could include the proximity guidance audio in the searching mode which was part of an initial prototype so that people can keep applying the idea of it for searching, but we left it out due to the experimental result from Ramôa et al. [80] showing that speech-based guidance was faster than sonar-based guidance and did not require training. However, with the addition of varying the pacing of the speech to inform distance toward the target and removal of the use of pitch from the sonar-based guidance in TADA, additional experiments might be needed to further compare the different ways of pinpointing a location using audio. We suspect that the differences might be marginal to small.

7.4 Similarities and Differences with Other Modalities and Formats of Access

Some interaction techniques supported by TADA align with how people are blind explore digital graphics with tactile sensations. Zhang et al. [99] found that there are five main exploration procedures. TADA achieves four of them. First, frame following refers to tracking the boundary of a graphic. This is inherently supported by TADA as the boundary of a diagram would be the boundary of the tablet. Second, surface sweeping is the back-and-forth movement to learn the features of objects. This is supported by the single-finger sweep interaction in TADA to learn the characteristics of the density and connectivity of diagram elements. Third, relative positioning is the back-and-forth movement between objects to obtain the relative locations, and this is utilized in the dwell + circle interaction to perceive the directions of connected links of a node, the space between them, and relatively how far they reach. This is also supported by the simultaneous interactions with TADA where people could interact with different parts of a diagram in parallel. Forth, absolute positioning is to understand the absolute locations of objects, and this is supported by all interactions in TADA as all diagram elements are represented in their absolute positions. The last procedure not supported is contour following, which is to trace the boundary of an object to perceive its size and shape, which are not inherent properties supported by TADA.

Tactile access still has its advantages. For example, tracking a link is natural in a tactile diagram, but in TADA, one might need to wiggle their finger as they navigate to stay on track. P2, P3, and P15 thought that using a tactile diagram would be easier. However, some participants thought otherwise. P20 said the navigation interactions in TADA served as an audio version of tactile access. P21 preferred audio-based accessibility solutions and personally thought TADA was better than a tactile diagram. P2 mentioned that TADA would be more portable than a tactile diagram, and P25 also mentioned using audio is more accessible whereas a tactile solution could be expensive. P18 commented TADA had the *"tactile nature"* because P18 could create a mental image of the diagram from the audio feedback.

Text descriptions are a common alternative way to access diagrams, sometimes providing multiple levels of detail despite being static. However, according to P3, TADA is a more intuitive and less tedious option. In addition to its other features, TADA also incorporates the use of alternative text as another means of accessing diagrams.

7.5 Making Diagram Data or Visualization Accessible

We refer to diagram data as a dataset containing data points and connectivity information between the data points and recognize that node-link visualization is one way to represent the data. We utilize the idea of a node-link diagram as the main representation in TADA due to its ability to draw attention to certain aspects of the data depending on the layout algorithm, and resemble semantic structure such as spatial locations in a map [56]. Visual diagrams can group relevant information together, support straightforward perceptual inferences [66], and support external memory [18]. Even if a diagram does not have meaningful spatial properties, it still exploits the properties of spatial relations to facilitate problem-solving [66]. We also reported the importance of diagrams and spatial information from the formative study we did with participants with visual impairments.

Information exists in represented forms of data, and TADA is another representation of the data. It is a nuanced question to determine if TADA is making the diagram data or the visualization of it accessible. Diagram data could already contain exact spatial locations using the node-link visualization, similar to how we created the sample and test diagrams for the evaluation study. Nevertheless, one can argue that the links are simply a visualization of the relations between data points which could be encoded audibly in some cases. However, TADA is not a direct translation of diagram visualization into an accessible version of it (DP6). It borrows important characteristics from a node-link representation, preserves the spatial data due to the aforementioned benefits of spatial relations, and represents the diagram data accordingly. It also adds new ways of exploring and understanding with audio and gesture while aiming to reduce the barrier to access non-visually. Various interaction techniques are designed to support tasks and gather different information from a diagram through audio feedback.

We believe that different representations exist for different types of data, and some are better than others for different purposes. This begs the question if there exists an audio-first representation of diagram data, and we recognize that this is an open research area.

7.6 Limitations of the Current Design of TADA and Future Improvements

TADA has limitations that need to be acknowledged and addressed as part of future work, related to its functionality, usability, usage in real-life situations, scalability, and adaptability.

7.6.1 Functionality Improvements

From the evaluation study, we noticed there were several modifications that could improve some of the currently supported features of TADA. We are in the process of incorporating these without altering the core functioning of other techniques. For example, TADA only filters for one type of node attribute now, and it could include a mechanism to support multiple layers of filters. The alt-text support can also include more levels of details, and both the alt-text and audio legend can add the ability to navigate through the sentences back and forth. The searching mode only supports one finger right now, and more could be supported with potentially clever ways to use audio to more quickly pinpoint multiple targets. This can also be combined with the proximity guidance audio which we currently left out to prioritize voice-based guidance in the search mode.

Multiple participants suggested TADA to support customizations. For example, P7 and P23 suggested the ability to slow down the audio summary from the five-finger dome interaction to hear it more in detail. P21 wanted to adjust the volume range of the proximity guidance audio, and P23 wanted to adjust the pitch range of sounds. P14 desired the option to turn certain audio off on demand to focus more on the other audio information. P7, P22, and P23 wanted to customize the speech synthesis voice and speed, and P9 suggested a verbosity setting so that a more advanced verbose level has fewer words.

More tasks could be supported such as annotation, and chunking [20]. Annotation refers to making notes on the diagram to offload information, and the notes could either become another attribute of a diagram element, or annotation could become a separate mode similar to accessing comments on a document. Chucking refers to the grouping of nodes into hierarchies.

In addition to diagram perception, the modification and creation of diagrams can also be investigated [38]. Furthermore, because TADA does not currently address the two lowest levels of diagram accessibility (not knowing the existence of a diagram and only knowing a placeholder of a diagram), the automatic extraction of diagram data is another area that could be explored using existing research methods [57, 60, 76] such as using image processing models. An example could be designing a way for TADA to receive a diagram from elsewhere, analyze it, and extract the underlying data to represent it for interactions.

7.6.2 Usability Improvements

We also observed during the evaluation study that the current design has some usability issues that could be addressed. For example, the pentagon shaped five-finger dome area may not fully cover the expected sub-region in a diagram, due to the boundary line formed between the thumb and pinky finger. To better cover a region, we could approximate the dome area as a hexagon with an additional vertex between the thumb and pinky finger, or approximate the area as an oval shape. Additionally, if a finger gets accidentally lifted off, one has to remove all five fingers and repeat the gesture all over again. Error prevention can be implemented in that if a finger gets lifted and then put down within a short period of time, the dome area stays. This could also happen due to input tracking errors, especially when two fingers stay too close to each other; they could be recognized as one single touch input. The density of nodes may also affect the distinguishability of the elements, as nodes may sound like they stick together when they are too close to each other. This could be improved by making the size of the nodes smaller if the speed of finger sweep is above a threshold so that the node audio sounds more discrete as nodes get interacted more quickly. Lastly, for people who find dwelling on a spot continuously to be awkward with other interactions with another finger, TADA can provide an option to allow people to lock and unlock the focus onto a particular diagram element without dwelling, similar to current interactions supported by VoiceOver or TalkBack. This would however require careful reconsideration of the potential impact on the current operations of the interface, such as the effect on multiple interactions happening simultaneously.

7.6.3 TADA in the Wild

TADA might encounter some challenges arising from its audio-based interface when operating with certain situational constraints. P13 mentioned that it would be hard to use a tablet while using a white cane. P5, P19, and P22 voiced their concerns for a noisy environment because the audio might not be clearly heard. Ahmetovic et al. [3] designed ways for screen readers to adapt intelligibly in noisy scenarios by adapting the volume, frequencies, and speed of audio information, and TADA can potentially borrow lessons from it to dynamically adapt to a noisy environment. P4 thought using TADA in real-life might not be quick enough and preferred a question-and-answer system for instant responses.

7.6.4 Scalability

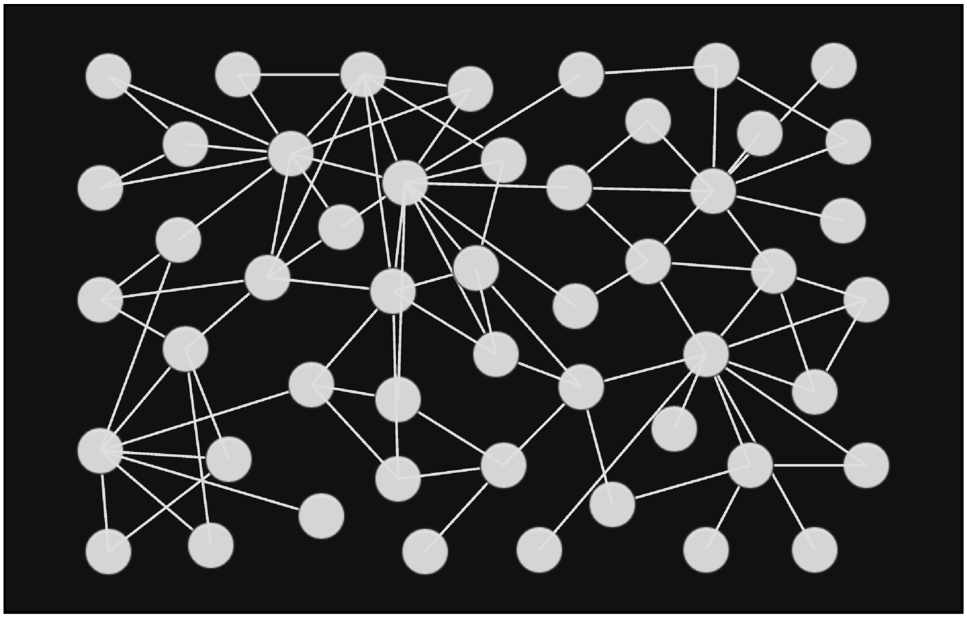
TADA's scalability regarding how much diagrammatic information can be represented in a single diagram is limited by the design of the touch interactions, the effectiveness of interaction techniques, and the utilization of sonic dimensions. Each node is represented by a circle with a certain area for a finger to interact with, and this area is expanded and large enough for a finger to be dwelt on to help prevent finger drifting. This could potentially cause a problem where multiple nodes exist in close proximity to each other. If overlap between nodes takes place, TADA is not yet designed to deal with it. Moreover, if a diagram contains spatial data, it also depends on the designer to craft a spaced-out diagram instead of one that is more difficult for touch interactions. Scalability is also partially affected by the use of locality during dwell + circle or radiate. If two nodes are connected and they are close to each other, the link will not be crossed when circling around it with two fingers that are not close enough to each other (which is uncomfortable in itself). In this case, the user can only infer from the proximity guidance sound that there is a link. If two links are too close to each other and almost parallel, it is also difficult to distinguish between them. P10 and P20 speculated that if a diagram is very complex, the audio feedback might not be helpful because it could contain too much information to be made sense of. Overall, the current design starts making interaction difficult at around 50 nodes (see Figure 7.1a), and becomes unusable at around 100 nodes (see Figure 7.1b). These limits can be extended through careful design, probably at the expense of some simplicity or usability. They also depend on the screen size of the hardware device and a larger screen can accommodate more information at a usable level (the device we used has a screen size of 10.6 inches). The most vulnerable interaction technique is likely dwell + circle / radiate due to the locality design, followed by the five-finger dome which currently has a 2-second duration to the audio summary. Increasing the size of the diagrams that are representable through a system like TADA is worthy of future study. Nevertheless, we believe that a large proportion of existing diagrams, such as most of those found in textbooks, instruction manuals and news media, fall below the threshold exemplified in Figure 7.1a.

More complex visualizations or alternative layouts could also be considered including features like directional links and curvy lines. Currently, the information represented by these either has to be encoded as attributes and interpreted by gaining the details with dwell + tap (resulting in more procedures to perform and time taken), and it has to be simplified to the basic layout with straight lines and equally sized nodes (potentially resulting in more overlapping between diagram elements). We need to think about the design of the information representation more carefully if we want to effectively include such complexities Filtering could help with more complex information from a diagram, but this is a subject that requires further investigation. An alternative to filtering is organizing a diagram in a hierarchical structure allowing people to start exploring from a higher hierarchy to help break down the information more. However, this would require radical changes to the nature of the interface. In particular, navigating hierarchies requires that the system retains a state about the current level and maintains the recognition of both higher and lower hierarchies.

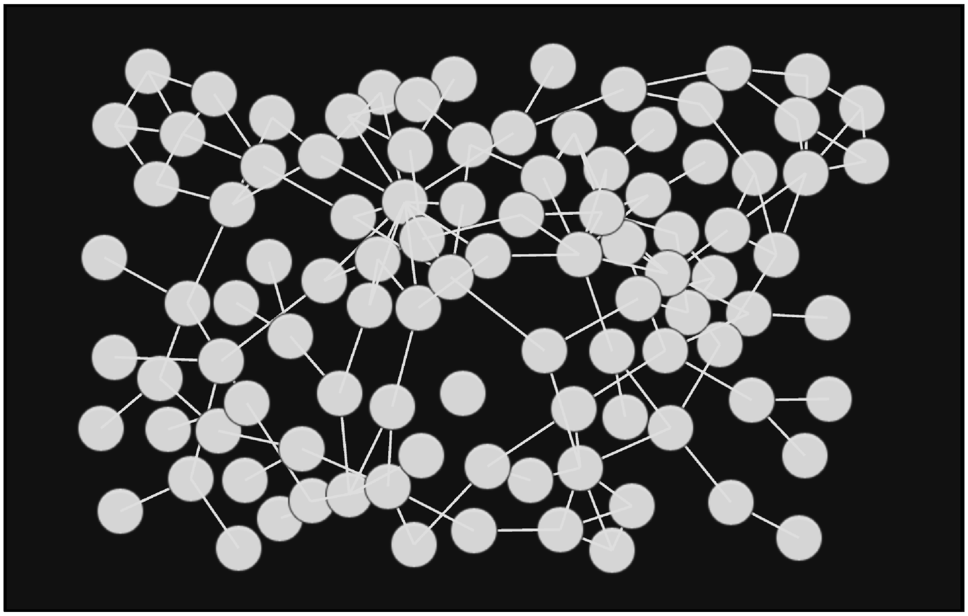
7.6.5 Adaptability

Participants expressed their interest in seeing the design of TADA to be applied to a phone form factor which is even more portable. This is an interesting extension of the design of TADA for it to be utilized in more contexts on the go, but it also offers less space for spatial perception because diagram elements are much closer to each other. Some interaction techniques will require adaptations to work through a smaller surface. P13 also suggested using a tablet for training on how to use TADA on a phone. With a phone, haptic support can be added to TADA as another modality. This needs additional considerations in terms of what information the haptic feedback can best represent.

Additionally, we could explore the possibility of applying TADA to other domains or tasks beyond node-link diagrams. For example, we could investigate how our system could support data analytics by enabling people to explore and analyze large relational datasets. Participants also expressed their curiosity about how TADA could potentially support other types of 2-dimensional diagrams or graphics such as geographical maps, indoor maps, webpages, statistical charts, role-playing game maps, and art pieces. TADA could also explore other diagram properties such as shapes and areas.



(a) A diagram with maximum complexity supported by TADA



(b) A diagram with too much complexity for TADA

Figure 7.1: Complex diagram examples

Chapter 8

Conclusion

We have explored the problem of diagram accessibility for people with visual impairments. We conducted a qualitative study with 15 participants to understand the experiences, challenges, and needs of people accessing diagrams. We proposed a ladder of diagram accessibility framework to capture the different levels of access that people can achieve. Based on our findings and insights, we designed and implemented TADA, a Touch-and-Audio-based Diagram Access system that enables people to interact with node-link diagrams on a tablet device. We evaluated TADA with 25 participants and found that it enabled people to perform tasks such as exploring, searching, and navigating node-link diagrams with high accuracy and low workload. People also gave positive feedback on the system's functionalities, usability, usefulness, and learnability. We discussed how TADA meets the design goals, the design approaches and trade-offs, comparisons with other accessible formats, limitations, and future work of our research. Our work contributes to the field of diagram accessibility by providing a thorough understanding of the problem, a novel system design, and an empirical evaluation of its effectiveness. We hope our work will encourage further research and development in the area of accessibility making information more accessible to people with visual impairments.

Bibliography

[1] Chuck: Unit generators reference (all).

[2] Anne Adams, Peter Lunt, and Paul Cairns. *A qualitative approach to HCI research*, page 138–157. Cambridge University Press, 1 edition, Aug 2008.

[3] Dragan Ahmetovic, Gabriele Galimberti, Federico Avanzini, Cristian Bernareggi, Luca Andrea Ludovico, Giorgio Presti, Gianluca Vasco, and Sergio Mascetti. Enhancing screen reader intelligibility in noisy environments. *IEEE Transactions on Human-Machine Systems*, page 1–10, 2023.

[4] Md Zubair Ibne Alam, Shehnaz Islam, and Enamul Hoque. Seechart: Enabling accessible visualizations through interactive natural language interface for people with visual impairments. Feb 2023. arXiv:2302.07742 [cs].

[5] Jérémy Albouys-Perrois, Jérémy Laviole, Carine Briant, and Anke M. Brock. Towards a multisensory augmented reality map for blind and low vision people: a participatory design approach. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, page 1–14, Montreal QC Canada, Apr 2018. ACM.

[6] R. Amar, J. Eagan, and John Stasko. *Low-Level Components of Analytic Activity in Information Visualization*, volume 15. Nov 2005.

[7] Android. Get started on android with talkback - android accessibility help.

[8] Apple. Control voiceover using the rotor on iphone.

[9] Apple. Turn on and practice voiceover on iphone.

[10] American Anthropological Association.

[11] Jack Atherton and Ge Wang. Chunity: Integrated audiovisual programming in unity. In *Proceedings of the 2018 conference on new interfaces for musical expression*, page 6, 2018.

[12] Sina Bahram. Multimodal eyes-free exploration of maps: Tikisi for maps. *ACM SIGACCESS Accessibility and Computing*, (106):3–11, Jun 2013.

[13] Suzanne P. Balik, Sean P. Mealin, Matthias F. Stallmann, Robert D. Rodman, Michelle L. Glatz, and Veronica J. Sigler. Including blind people in computing through access to graphs. In *Proceedings of the 16th international ACM SIGACCESS conference on Computers and accessibility*, ASSETS '14, page 91–98, New York, NY, USA, Oct 2014. Association for Computing Machinery.

[14] David Barter and Peter Coppin. A diagram must never be ten thousand words: Text-based (sentential) approaches to diagrams accessibility limit users' potential for normative agency. In Valeria Giardino, Sven Linker, Richard Burns, Francesco Bellucci, Jean-Michel Boucheix, and Petrucio Viana, editors, *Diagrammatic Representation and Inference*, Lecture Notes in Computer Science, page 86–93, Cham, 2022. Springer International Publishing. Citation Key: barter-DiagramMustNever-2022.

[15] P. Blenkhorn and D. G. Evans. Using speech and touch to enable blind people to access schematic diagrams. *Journal of Network and Computer Applications*, 21(1):17–29, Jan 1998. Citation Key: blenkhorn-UsingSpeechTouch-1998.

[16] Ulrik Brandes, Markus Eiglsperger, Jürgen Lerner, and Christian Pich. *Graph Markup Language (GraphML).* Discrete mathematics and its applications. CRC Press, Taylor & Francis Group, Boca Raton, 2010.

[17] Anke Brock and Christophe Jouffrais. Interactive audio-tactile maps for visually impaired people. *ACM SIGACCESS Accessibility and Computing*, (113):3–12, Nov 2015.

[18] A. Brown, R. Stevens, and S. Pettifer. Issues in the non-visual presentation of graph based diagrams. In *Proceedings. Eighth International Conference on Information Visualisation, 2004. IV 2004.*, page 671–676, Jul 2004.

[19] Andy Brown, Steve Pettifer, and Robert Stevens. Evaluation of a non-visual molecule browser. In *Proceedings of the 6th international ACM SIGACCESS con-*

ference on Computers and accessibility, Assets '04, page 40–47, New York, NY, USA, Sep 2003. Association for Computing Machinery.

[20] Andy Brown, Robert Stevens, and Steve Pettifer. Making graph-based diagrams work in sound: The role of annotation. *Human–Computer Interaction*, 28(3):193–221, May 2013.

[21] Emeline Brule, Gilles Bailly, Anke Brock, Frederic Valentin, Grégoire Denis, and Christophe Jouffrais. Mapsense: Multi-sensory interactive maps for children living with visual impairments. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, page 445–457, San Jose California USA, May 2016. ACM.

[22] Matt Calder, Robert F. Cohen, Jessica Lanzoni, and Yun Xu. Plumb: an interface for users who are blind to display, create, and modify graphs. In *Proceedings of the 8th international ACM SIGACCESS conference on Computers and accessibility*, Assets '06, page 263–264, New York, NY, USA, Oct 2006. Association for Computing Machinery.

[23] Dustin Carroll, Suranjan Chakraborty, and Jonathan Lazar. *Designing Accessible Visualizations: The Case of Designing a Weather Map for Blind Users*, volume 8009 of *Lecture Notes in Computer Science*, page 436–445. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[24] DIAGRAM Center.

[25] Jinho Choi, Sanghun Jung, Deok Gun Park, Jaegul Choo, and Niklas Elmqvist. Visualizing for the non-visual: Enabling the visually impaired to use visualization. *Computer Graphics Forum*, 38(3):249–260, 2019. Citation Key: choi-VisualizingNonVisualEnabling-2019a.

[26] Robert F. Cohen, Rui Yu, Arthur Meacham, and Joelle Skaff. Plumb: displaying graphs to the blind using an active auditory interface. In *Proceedings of the 7th international ACM SIGACCESS conference on Computers and accessibility*, Assets '05, page 182–183, New York, NY, USA, Oct 2005. Association for Computing Machinery. Citation Key: cohen-PLUMBDisplayingGraphs-2005.

[27] Perry R Cook and Gary P Scavone. The synthesis toolkit (stk). 1999.

[28] Peter Coppin. *Perceptual-cognitive Properties of Pictures, Diagrams, and Sentences: Toward a Science of Visual Information Design.* Thesis, Mar 2014. Accepted: 2014-03-27T18:54:58Z.

[29] Peter Coppin, Abrose Li, and Michael Carnevale. Iconic properties are lost when translating visual graphics to text for accessibility. *Cognitive Semiotics*, 2016.

[30] Charlie Cross, Deniz Cetinkaya, and Huseyin Dogan. Transforming diagrams' semantics to text for visually impaired. In Aaron Marcus and Elizabeth Rosenzweig, editors, *Design, User Experience, and Usability. Interaction Design*, Lecture Notes in Computer Science, page 339–350, Cham, 2020. Springer InternaTheretional Publishing. Citation Key: cross-TransformingDiagramsSemantics-2020.

[31] Julie Ducasse, Anke M. Brock, and Christophe Jouffrais. *Accessible Interactive Maps for Visually Impaired Users*, page 537–584. Springer International Publishing, Cham, 2018.

[32] Emory J. Edwards, Michael Gilbert, Emily Blank, and Stacy M. Branham. How the alt text gets made: What roles and processes of alt text creation can teach us about inclusive imagery: What roles and processes of alt text creation can teach us about inclusive imagery. *ACM Transactions on Accessible Computing*, page 3587469, Mar 2023.

[33] Be My Eyes. Be my eyes - see the world together.

[34] Danyang Fan, Alexa Fay Siu, Wing-Sum Adrienne Law, Raymond Ruihong Zhen, Sile O'Modhrain, and Sean Follmer. Slide-tone and tilt-tone: 1-dof haptic techniques for conveying shape characteristics of graphs to blind users. In *CHI Conference on Human Factors in Computing Systems*, page 1–19, New Orleans LA USA, Apr 2022. ACM. Citation Key: fan-SlideToneTiltTone1DOF-2022b.

[35] Martin H. Frické. *Data-Information-Knowledge-Wisdom (DIKW) Pyramid, Framework, Continuum*, page 1–4. Springer International Publishing, Cham, 2018.

[36] Thomas M. J. Fruchterman and Edward M. Reingold. Graph drawing by force-directed placement. *Software: Practice and Experience*, 21(11):1129–1164, 1991.

[37] Cagatay Goncu and Kim Marriott. *GraVVITAS: Generic Multi-touch Presentation of Accessible Graphics*, volume 6946 of *Lecture Notes in Computer Science*, page 30–48. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[38] William Grussenmeyer and Eelke Folmer. Audiodraw: user preferences in non-visual diagram drawing for touchscreens. In *Proceedings of the 13th International Web for All Conference*, page 1–8, Montreal Canada, Apr 2016. ACM. Citation Key: grussenmeyer-AudioDrawUserPreferences-2016.

[39] Richa Gupta, M. Balakrishnan, and P.V.M. Rao. Tactile diagrams for the visually impaired. *IEEE Potentials*, 36(1):14–18, Jan 2017. Citation Key: gupta-TactileDiagramsVisually-2017.

[40] Sandra G. Hart. Nasa-task load index (nasa-tlx); 20 years later. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 50(9):904–908, Oct 2006.

[41] Sandra G. Hart and Lowell E. Staveland. *Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research*, volume 52 of *Human Mental Workload*, page 139–183. North-Holland, Jan 1988.

[42] Liang He, Zijian Wan, Leah Findlater, and Jon E. Froehlich. Tactile: A preliminary toolchain for creating accessible graphics with 3d-printed overlays and auditory annotations. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*, page 397–398, Baltimore Maryland USA, Oct 2017. ACM. Citation Key: he-TacTILEPreliminaryToolchain-2017.

[43] Marti A Hearst. Show it or tell it? text, visualization, and their combination. 2023.

[44] Leona M Holloway, Cagatay Goncu, Alon Ilsar, Matthew Butler, and Kim Marriott. Infosonics: Accessible infographics for people who are blind using sonification and voice. In *CHI Conference on Human Factors in Computing Systems*, page 1–13, New Orleans LA USA, Apr 2022. ACM. Citation Key: holloway-InfosonicsAccessibleInfographics-2022.

[45] Anett Hoppe, David Morris, and Ralph Ewerth. Evaluation of automated image descriptions for visually impaired students. In Ido Roll, Danielle McNamara, Sergey Sosnovsky, Rose Luckin, and Vania Dimitrova, editors, *Artificial Intelligence in Education*, Lecture Notes in Computer Science, page 196–201, Cham, 2021. Springer International Publishing. Citation Key: hoppe-EvaluationAutomatedImage-2021.

[46] M. Horstmann, M. Lorenz, A. Watkowski, G. Ioannidis, O. Herzog, A. King, D. G. Evans, C. Hagen, C. Schlieder, A.-M. Burn, N. King, H. Petrie, S. Dijkstra, and

D. Crombie. Automated interpretation and accessible presentation of technical diagrams for blind people. *New Review of Hypermedia and Multimedia*, 10(2):141–163, Dec 2004. Citation Key: horstmann-AutomatedInterpretationAccessible-2004.

[47] W3C Web Accessibility Initiative (WAI). Tips and tricks.

[48] j1mmyto9. Speed to text in unity ios use native speech recognition.

[49] Shakila Cherise S Joyner, Amalia Riegelhuth, Kathleen Garrity, Yea-Seul Kim, and Nam Wook Kim. Visualization accessibility in the wild: Challenges faced by visualization designers. In *CHI Conference on Human Factors in Computing Systems*, page 1–19, New Orleans LA USA, Apr 2022. ACM. Citation Key: joyner-VisualizationAccessibilityWild-2022.

[50] Crescentia Jung, Shubham Mehta, Atharva Kulkarni, Yuhang Zhao, and Yea-Seul Kim. Communicating visualizations without visuals: Investigation of visualization alternative text for people with visual impairments. *IEEE Transactions on Visualization and Computer Graphics*, 28(1):1095–1105, Jan 2022. Citation Key: jung-CommunicatingVisualizationsVisuals-2022.

[51] Nikolaos Kaklanis, Konstantinos Votis, and Dimitrios Tzovaras. A mobile interactive maps application for a visually impaired audience. In *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility*, W4A '13, page 1–2, New York, NY, USA, May 2013. Association for Computing Machinery.

[52] Jyoti Kala, Kunal Kwatra, Rizul Kumar, Piyush Chanana, P V M Rao, and M Balakrishnan. Tactile diagrams for science and mathematics: Design, production and experiences of students with blindness. 2017.

[53] Shaun K. Kane, Jeffrey P. Bigham, and Jacob O. Wobbrock. Slide rule: making mobile touch screens accessible to blind people using multi-touch interaction techniques. In *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility - Assets '08*, page 73, Halifax, Nova Scotia, Canada, 2008. ACM Press.

[54] Shaun K. Kane, Meredith Ringel Morris, Annuska Z. Perkins, Daniel Wigdor, Richard E. Ladner, and Jacob O. Wobbrock. Access overlays: improving non-visual access to large touch screens for blind users. In *Proceedings of the 24th an-*

nual ACM symposium on User interface software and technology, UIST '11, page 273–282, New York, NY, USA, Oct 2011. Association for Computing Machinery.

[55] Jiro Kawakita. The original kj method. *Tokyo: Kawakita Research Institute*, 5:1991, 1991.

[56] René Keller, Claudia M. Eckert, and P. John Clarkson. Matrices or node-link diagrams: Which visual representation is better for visualising connectivity models? *Information Visualization*, 5(1):62–76, Mar 2006. Citation Key: keller-MatricesNodeLinkDiagrams-2006.

[57] Aniruddha Kembhavi, Mike Salvato, Eric Kolve, Minjoon Seo, Hannaneh Hajishirzi, and Ali Farhadi. A diagram is worth a dozen images. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision – ECCV 2016*, Lecture Notes in Computer Science, page 235–251, Cham, 2016. Springer International Publishing. Citation Key: kembhavi-DiagramWorthDozen-2016.

[58] Andrea R. Kennel. Audiograf: a diagram-reader for the blind. In *Proceedings of the second annual ACM conference on Assistive technologies*, Assets '96, page 51–56, New York, NY, USA, Apr 1996. Association for Computing Machinery.

[59] Prerna Khanna, Shirin Feiz, and Jian Xu. Accesswear: Making smartphone applications accessible to blind users. 2023.

[60] Daesik Kim, YoungJoon Yoo, Jeesoo Kim, Sangkuk Lee, and Nojun Kwak. Dynamic graph generation network: Generating relational knowledge from diagrams. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, page 4167–4175, Salt Lake City, UT, Jun 2018. IEEE. Citation Key: kim-DynamicGraphGeneration-2018.

[61] Gyeongri Kim, Jiho Kim, and Yea-Seul Kim. "explain what a treemap is": Exploratory investigation of strategies for explaining unfamiliar chart to blind and low vision users. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, page 1–13, Hamburg Germany, Apr 2023. ACM.

[62] Jiho Kim. Exploring chart question answering for blind and low vision users. 2023.

[63] N. W. Kim, S. C. Joyner, A. Riegelhuth, and Y. Kim. Accessible visualization: Design space, opportunities, and challenges. *Computer Graphics Forum*, 40(3):173–188, Jun 2021. Citation Key: kim-AccessibleVisualizationDesign-2021a.

[64] Alasdair King, Paul Blenkhorn, David Crombie, Sijo Dijkstra, Gareth Evans, and John Wood. Presenting uml software engineering diagrams to blind people. In Klaus Miesenberger, Joachim Klaus, Wolfgang L. Zagler, and Dominique Burger, editors, *Computers Helping People with Special Needs*, Lecture Notes in Computer Science, page 522–529, Berlin, Heidelberg, 2004. Springer. Citation Key: king-PresentingUMLSoftware-2004.

[65] Jenna L. Gorlewicz, Jennifer L. Tennison, Hari P. Palani, and Nicholas A. Giudice. *The Graphical Access Challenge for People with Visual Impairments: Positions and Pathways Forward*. IntechOpen, Sep 2019.

[66] Jill H. Larkin and Herbert A. Simon. Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11(1):65–100, 1987.

[67] Jonathan Lazar, Suranjan Chakraborty, Dustin Carroll, Robert Weir, Bryan Sizemore, and Haley Henderson. Development and evaluation of two prototypes for providing weather map data to blind users through sonification. 8(4), 2013.

[68] Bongshin Lee, Catherine Plaisant, Cynthia Sims Parr, Jean-Daniel Fekete, and Nathalie Henry. Task taxonomy for graph visualization. In *Proceedings of the 2006 AVI workshop on BEyond time and errors: novel evaluation methods for information visualization*, BELIV '06, page 1–5, New York, NY, USA, May 2006. Association for Computing Machinery.

[69] Lumivero. Nvivo - lumivero.

[70] Michał Maćkowski, Piotr Brzoza, Mateusz Kawulok, Rafał Meisel, and Dominik Spinczyk. Multimodal presentation of interactive audio-tactile graphics supporting the perception of visual information by blind people. *ACM Transactions on Multimedia Computing, Communications, and Applications*, page 3586076, Mar 2023. Citation Key: mackowski-MultimodalPresentationInteractive-2023.

[71] David Mcgookin, Stephen Brewster, and WeiWei Jiang. Investigating touchscreen accessibility for people with visual impairments. volume 358, Oct 2008.

[72] Oussama Metatla, Nick Bryan-Kinns, and Tony Stockman. Using hierarchies to support non-visual access to relational diagrams. volume 1, page 215–225, Jan 2007.

[73] Prerna Mishra, Santosh Kumar, Mithilesh Kumar Chaube, and Urmila Shrawankar. Chartvi: Charts summarizer for visually impaired. *Journal of Computer Languages*, 69:101107, Apr 2022.

[74] Vishnu Nair, Hanxiu "Hazel" Zhu, and Brian A. Smith. Imageassist: Tools for enhancing touchscreen-based image exploration systems for blind and low vision users. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, page 1–17, Apr 2023. arXiv:2302.09124 [cs].

[75] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69(2):026113, Feb 2004. arXiv:cond-mat/0308217.

[76] OpenAI. Gpt-4 technical report. (arXiv:2303.08774), Mar 2023. arXiv:2303.08774 [cs].

[77] Helen Petrie, Neil King, Anne-Marie Burn, and Peter Pavan. Providing interactive access to architectural floorplans for blind people. *British Journal of Visual Impairment*, 24(1):4–11, Jan 2006.

[78] Helen Petrie, Christoph Schlieder, Paul Blenkhorn, Gareth Evans, Alasdair King, Anne-Marie O'Neill, George T. Ioannidis, Blaithin Gallagher, David Crombie, Rolf Mager, and Maurizio Alafaci. Tedub: A system for presenting and exploring technical drawings for blind people. In Klaus Miesenberger, Joachim Klaus, and Wolfgang Zagler, editors, *Computers Helping People with Special Needs*, Lecture Notes in Computer Science, page 537–539, Berlin, Heidelberg, 2002. Springer.

[79] Benjamin Poppinga, Charlotte Magnusson, Martin Pielot, and Kirsten Rassmus-Gröhn. Touchover map: audio-tactile exploration of interactive maps. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, page 545–550, Stockholm Sweden, Aug 2011. ACM.

[80] Gaspar Ramôa, Vincent Schmidt, and Peter König. Developing dynamic audio navigation uis to pinpoint elements in tactile graphics. *Multimodal Technologies and Interaction*, 6(12):113, Dec 2022. Citation Key: ramoa-DevelopingDynamicAudio-2022.

[81] Jennifer Rowley. The wisdom hierarchy: representations of the dikw hierarchy. *Journal of Information Science*, 33(2):163–180, Apr 2007.

[82] Bahador Saket, Paolo Simonetto, Stephen Kobourov, and Katy Börner. Node, node-link, and node-link-group diagrams: An evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2231–2240, Dec 2014.

[83] Gary P Scavone and Perry R Cook. Rtmidi, rtaudio, and a synthesis toolkit (stk) update. 2005.

[84] Mathieu Simonnet, Cécile Bothorel, Luiz Felipe Maximiano, and André Thépaut. Geotablet, une application cartographique pour les personnes déficientes visuelles. 2012.

[85] Jing Su, Alyssa Rosenzweig, Ashvin Goel, Eyal De Lara, and Khai N. Truong. Timbremap: enabling the visually-impaired to use maps on touch-enabled devices. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*, page 17–26, Lisbon Portugal, Sep 2010. ACM.

[86] Saiganesh Swaminathan, Thijs Roumen, Robert Kovacs, David Stangl, Stefanie Mueller, and Patrick Baudisch. Linespace: A sensemaking platform for the blind. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, page 2175–2185, San Jose California USA, May 2016. ACM.

[87] Benny J Tang, Angie Boggust, and Arvind Satyanarayan. Vistext: A benchmark for semantically rich chart captioning.

[88] Unity Technologies. Maximize multiplatform game development — unity.

[89] Jennifer L. Tennison and Jenna L. Gorlewicz. Non-visual perception of lines on a multimodal touchscreen tablet. *ACM Transactions on Applied Perception*, 16(1):1–19, Jan 2019.

[90] John R Thompson, Jesse J Martinez, Alper Sarikaya, Edward Cutrell, and Bongshin Lee. Chart reader: Accessible visualization experiences designed with screen reader users. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, page 1–18, New York, NY, USA, Apr 2023. Association for Computing Machinery.

[91] Márcio Josué Ramos Torres and Regina Barwaldt. Approaches for diagrams accessibility for blind people: a systematic review. In *2019 IEEE Frontiers in Education Conference (FIE)*, page 1–7, Oct 2019. Citation Key: torres-ApproachesDiagramsAccessibility-2019.

[92] Unity Technologies. Unity real-time development platform.

[93] Ge Wang and Perry R Cook. Chuck: A concurrent, on-the-fly, audio programming language. In *Proceedings of the 2003 International Computer Music Conference*, page 8, 2003.

[94] R. Wang, C. Jung, and Y. Kim. Seeing through sounds: Mapping auditory dimensions to data and charts for people with visual impairments. *Computer Graphics Forum*, 41(3):71–83, 2022. Citation Key: wang-SeeingSoundsMapping-2022.

[95] Colin Ware. *Information Visualization: Perception for Design.* Elsevier, 2013. Google-Books-ID: qFmS95vf6H8C.

[96] Naohide Yamamoto and Amy L. Shelton. Visual and proprioceptive representations in spatial memory. *Memory & Cognition*, 33(1):140–150, Jan 2005.

[97] Yalong Yang, Kim Marriott, Matthew Butler, Cagatay Goncu, and Leona Holloway. Tactile presentation of network data: Text, matrix or diagram? In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, page 1–12, New York, NY, USA, Apr 2020. Association for Computing Machinery. Citation Key: yang-TactilePresentationNetwork-2020.

[98] yWorks. yed live.

[99] Ting Zhang, Bradley S. Duerstock, and Juan P. Wachs. Classification of blind users' image exploratory behaviors using spiking neural networks. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(4):1032–1041, Apr 2020. Citation Key: zhang-ClassificationBlindUsers-2020.

[100] Zhuohao Zhang and Jacob O. Wobbrock. A11yboard: Using multimodal input and output to make digital artboards accessible to blind users. In *Adjunct Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, UIST '22 Adjunct, page 1–4, New York, NY, USA, Oct 2022. Association for Computing Machinery.

[101] Haixia Zhao, Catherine Plaisant, Ben Shneiderman, and Jonathan Lazar. Data sonification for users with visual impairment: A case study with georeferenced data. *ACM Transactions on Computer-Human Interaction*, 15(1):4:1–4:28, May 2008.

Appendix A

Formative Study Procedure

Semi-structured interview

Demographics questions

- What is your age?
- What is your gender?
- What is your professional background?
- What is your degree of vision loss?
- When was your onset of vision loss?
- Do you access diagrammatic information for day-to-day activities at work or for personal tasks? If so, how often?

Participant questions

The interviewer asks interviewees to choose two or three scenarios where they encounter diagrammatic information. Interviewees can screen-share samples of their diagrammatic information if possible. For each scenario:

1. Describe briefly the diagrammatic information.
2. Describe briefly the context of the encounter in terms of when, where, how, who, and why.
3. To which degree were you successful in accessing the diagrammatic information? (0 - not successful at all, I had to give up; to 5 - perfectly successful, I got everything I needed from it very easily)
4. How often do you encounter the scenario?
5. What were the motives for or objectives of accessing the information?
6. What is the importance of the information?
7. Was it an individual, collaborative, or hybrid setting?
8. What was the surrounding physical environment? For example, was it in public or private? Was it personal or professional?
9. What media was the information presented in? Was it printed, electronic, verbal, or others?
10. What did you do when you accessed the information? What tools did you use to access the information, if any?
11. What types of tasks do you do with the diagrammatic information? Which are important to this scenario, if any?

    (The followings are some examples: Gaining a general idea of the diagram; Searching for a piece of information [e.g., a node, or information on a node]; Exploring the diagram in detail; Browsing or navigating through the diagram; Annotating the diagram; Other tasks [Interviewer asks interviewees to specify what tasks])

12. What went well and smoothly when trying to access the diagram?

13. What are the main challenges that you encountered when trying to access the diagram?
    (Interviewer could provide some ideas if they do not come up with them on their own; e.g., Did you have to do something inconvenient? Did it take a long time to start the tool? Was the circumstance of use awkward?)
14. How did you try to address these challenges, if at all?
15. Do you have any ideas that would help you access diagrammatic information in better ways?
    (Interviewer could afterwards ask more specific questions about the importance and the usefulness of spatial information, types of haptic / tactile feedback and sounds that they would utilize, and physical objects that they are familiar with)

Appendix B

Evaluation Study Procedure

Study Protocol

1. (3 min) Researcher asks participant about their background information.

1.1. What is your age?

1.2. What is your gender?

1.3. What is your educational and professional background?

1.4. What is your degree of vision loss?

1.5. When was your onset of vision loss?

1.6. How much does vision loss impact your life?

1.7. How frequently do you access diagram information in your daily life? (eg. on a weekly basis)

1.8. Briefly explain how do you currently access diagrams in your daily life?

2. (2 min) Researcher explains the general context of the user study, including the main goal of the prototype, what will be demonstrated, and a brief description of a sample diagram. Researcher answers any questions from participant.

Script: "Thank you for participating in this evaluation session of our prototype which aims to make diagrams accessible to individuals with vision loss. I will demonstrate five main interaction modes, and for each demonstration, you will have the chance to try it out, and answer some questions from me while thinking aloud. The prototype is built on a tablet which is just in front of you. It has a diagram on the screen like a flowchart, transportation lines, or in our case a friendship diagram – Imagine it shows a map of where people are in a room, and if there is a connection line between two people, it means they are friends. A diagram consists of items and connections between them. Items refer to the entities or objects such as people, and the connections are relations between the items. To access the diagram, the prototype will produce audio depending on how you interact with it using your hands and fingers. During the demonstration, there will be tactile components put onto the tablet for you to get familiar with the prototype and its interaction modes."

"We will begin the training demonstrations. Do you mind if I could touch and hold your fingers to help you move around and get familiar with the gestures? "

3. (12 mins total) (2 min) Researcher explains and demonstrates the single-finger mode and basic sounds of nodes and links.

Script: "To demonstrate the single-finger mode, I will put my finger onto the screen and begin moving it. As I move it, you can hear there is a brass-like sound, and there is a string-like pluck sound. The brass sound means it is a node, and the pluck sound means it is a link. A higher pitched brass means the node is more connected. A higher pitched pluck means the link is shorter."

3.1. (2 min) Participant is asked to try out the single-finger mode. A tactile overlay is used to create a tactile modal of the diagram to help participant understand how the interaction mode with diagram works. This is then removed.

3.2. (1 min) Repeat above without overlay.

3.3. Participant is asked to interact with the nodes utilizing single-finger mode. Researcher asks question:

3.3.1. (3 min) Could you describe the overall layout of the diagram?

3.3.2. (1 min)

A) Which part of the diagram has the most connections between people?

B) Which part has the least?

3.3.3. (1 min)

A) Which part of the diagram has the most people?

B) Which part has the least?

3.3.4. (2 min) What do you think of this mode? Why does it work / not work for you?

4. (14 mins total) (2.5 min) Researcher explains and demonstrates the five-finger mode and the associated audio.

Script: “Next, the five-finger mode is meant for us to access only a partial area of the diagram. What I do is to put down my 5 fingers and this creates an area that is almost circular on the diagram, and we hear the sounds of nodes and links only from this region of the diagram. The order of the sounds goes from the first node, its connected links, then move on to the next nodes. The durations for each set of a node and its connected links are the same. So the pacing is important in that the more things we have under our hand, the quicker they would sound. The ‘ding’ sound signifies the start of a repetition of the sequence of node and link sounds. And I can always change the size of this area by expanding or contract my fingers.”

4.1. (2 min) Participant is asked to try out the five-finger mode with a tactile overlay to help them get familiar with this mode.

4.2. (2 min) Repeat above without overlay.

4.3. Participant is asked to utilize the five-finger mode to interact with the diagram and compare the 4 quadrants of the diagram. Researcher asks questions:

4.3.1. (2.5 min)

A) Which quadrant of the diagram has the most connections between people?

B) Which part the least?

4.3.2. (1 min)

A) Which quadrant of the diagram has the most people?

B) Which has the least?

4.3.3. (1 min) Which quadrant has people with unequal numbers of friends?

4.3.4. (2 min) What do you think of this mode? Why does it work / not work for you?

5. (4 mins total) (1 min) Researcher explains and demonstrates secondary tap mode and the associated audio.

Script: “Now to understand what the nodes and links mean contextually, we hold our first finger on a node or link, then use a second finger to do a quick tap near the first finger to hear the properties embedded within the node or link to learn their details. “

5.1. (1 min) Participant is asked to try out the secondary tap mode.

5.2. Researcher asks question:

5.2.1. (1 min) Given a person node, what is the name and professional background associated with it?

5.2.2. (1 min) What do you think of this mode? Why does it work / not work for you?

6. (5 mins total) (1.5 min) Researcher explains and demonstrates search mode and the associated audio.

Script: “We could also do a search by speaking what we want to the system, and this is initialized by a downward flick. Once the system starts to listen for our input, we can tell it what you want, and put a finger onto the screen to start searching. We hear directional words informing us where to move our finger towards the desired direction. When we finally reach the target, the system plays a triumph-like audio signifying we have found what we were searching for. ”

6.1. (1.5 min) Participant is asked to try out the search mode.

6.2. Researcher asks questions:

6.2.1. (1 min) Where is Bob? Locate Bob.

6.2.2. (1 min) What do you think of this mode? Why does it work / not work for you?

7. (7 mins total) (2 min) Researcher explains and demonstrates find-relations mode and the associated audio.

Script: “We can find the relations, or the connected links of a node. To do this, we keep one finger on a node. We put down a second finger near this node, and we can move the second finger around the first finger on the node in a circular fashion to hear the connected links. There is an underlying audio with varying volume, the louder it is, the nearer we are to a link.”

7.1. (1.5 min) Participant is asked to try out the find-relations mode with a tactile overlay.

7.2. (1.5 min) Repeat the above without an overlay.

7.3. Researcher asks questions:

7.3.1. (1 min) How many friends does Bob have?

7.3.2. (1 min) What do you think of this mode? Why does it work / not work for you?

8. (8 mins total) (2 min) Researcher explains and demonstrates navigating mode and the associated audio.

Script: “We can also navigate from a node to its neighbour. Following the previous mode, after selecting a link, we can then move our second finger towards the direction of the link and towards the target node. We can wiggle our finger to stay on track. The underlying audio works in the same way, the louder it is, the nearer we are to the target node. When we finally reach, a triumph-like audio is played. ”

8.1. (1.5 min) Participant is asked to try out the navigating mode with a tactile overlay.

8.2. (1.5 min) Repeat the above without an overlay.

8.2. Researcher asks questions:

8.2.1. (2 min) Navigate from Bob to any of his friends.

8.2.2. (1 min) What do you think of this mode? Why does it work / not work for you?

9. (4 min) Lastly, researcher would ask participant questions about the experience of using the prototype following the NASA-TLX questionnaire, and any suggestions for improvement. Lower score is better.

9.1. Mental demand: How mentally demanding was the task?

9.2. Physical demand: How physically demanding was the task?

9.3. Temporal demand: How hurries or rushed was the pace of the task?

9.4. Performance: How successful were you in accomplishing what you were asked to do? (perfect to failure)

9.5. Effort: How hard did you have to work to accomplish your level of performance?

9.6. Frustration: How insecure, discourage, irritated, stressed, and annoyed were you?