ADVANCING OMNIMODALITY: EXPANDING HUMAN CREATIVITY THROUGH ADAPTABLE AND ACCESSIBLE MULTIMODAL COMPUTING SYSTEMS

Joshua Urban Davis
josh.u.davis.gr@dartmouth.edu

Follow this and additional works at: https://digitalcommons.dartmouth.edu/dissertations

Part of the Computational Engineering Commons

Recommended Citation
https://digitalcommons.dartmouth.edu/dissertations/233

This Thesis (Ph.D.) is brought to you for free and open access by the Theses and Dissertations at Dartmouth Digital Commons. It has been accepted for inclusion in Dartmouth College Ph.D Dissertations by an authorized administrator of Dartmouth Digital Commons. For more information, please contact dartmouthdigitalcommons@groups.dartmouth.edu.
ADVANCING OMNIMODALITY: EXPANDING HUMAN CREATIVITY THROUGH ADAPTABLE AND ACCESSIBLE MULTIMODAL COMPUTING SYSTEMS

A Thesis
Submitted to the Faculty
in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy

in
Computer Sciences
by Josh Urban Davis
Guarini School of Graduate and Advanced Studies
Dartmouth College
Hanover, New Hampshire
April 2024

Examining Committee:

____________________________________
Xing-Dong Yang, Chair

____________________________________
Devin Balkcom

____________________________________
Elizabeth Murnane

____________________________________
Paul Asente (Adobe Research)

F. Jon Kull, Ph.D.
Dean of the Guarini School of Graduate and Advanced Studies
Abstract

Emerging technologies have given us a whole host of new ways for people to be creative. From the immersive worlds of AR/VR to the synthesis powers of large-language models and generative AI, these new tools hold the potential to reshape human expression and creativity. But how can we ensure that these new modalities are accessible to everyone, even those who aren’t able bodied? This thesis advocates for a human-centered approach to the development of many-modal systems. I will probe how our machines support, direct, or inhibit creativity as a mode of problem solving through 6 novel multimodal prototype interfaces and discuss how these systems support creativity for people who may otherwise be excluded. We will explore how multimodal systems allow people to interact according to their preferences and abilities, and how this flexibility creates space between modalities for accessible, rich, and creative computing. This includes discussing what we can learn about creativity by engaging people with disabilities in co-design, and how supporting this model of problem solving is imperative for designing future system interfaces that are inclusive and usable by all people.
Preface

Something as daunting and ambitious as a doctoral thesis never happens alone, but is made possible by the support of innumerable generous folks who lend their minds, advice, and support. Throughout this journey I’ve been helped by so many people it would be impossible to name them all here. Still, I will try my best.

I am extremely thankful to my committee members Devin Balkcom and Elizabeth Murnane, and long-time mentor and external appraiser Paul Asente for their constructive feedback and guidance helping me structure and compile this work. I would also like to thank my advisor, Xing-Dong Yang. I am also thankful to Autodesk Research, Microsoft Research, and Adobe Research for hosting me for internships, and to my mentors there including Fraser Anderson, Justin Matejka, George Fitzmaurice, Tovi Grossman, Ed Cutrell, John Tang, Merideth Ringel Morris, Teddy Seyed, Aaron Hertzman, Rubaiat Habib, and Nathan Carr.

I am incredibly grateful to all of my coauthors, many of whom have become long-term mentors and repeat collaborators, including Parmit Chilana, Miriam Sturdee, Makayla Lewis, Jun Gong, Johann Wentzel, Mafalda Gamboa, Sarah Fdili Alaoui, Claire Elisabeth Ohlenschlager, William Gaver, Eli Blevis, Lian Loke, James Mahoney, Te-yen Wu, Emily Whiting, Angelika Strohmayer, Katta Spiel, Nantia Koulidou, Kelly Mack, Maitraye Das, Dhruv Jain, Danielle Bragg, Andrew Begel, Erin Beneteau, John Miers, Abraham Glasser, Joon Sung Park, and Venkatesh Potluri. Thank you.
For their invaluable help with TangibleCircuits, I would like to thank Josh Miele, Carol Center for the Blind, and LightHouse San Francisco for all of their support in this work. Clio began as a project during an internship with Adobe Research, and I would like to thank Stephen DiVerdi and Wil Li for their guidance, support, and encouragement. For their help and guidance with Erato, I would like to thank Lessley Hernandez, Kay Frei, and Seyed Hussaini for their help and enthusiasm with this work, as well as Alison May and the Dartmouth Student Accessibility Services for their support and encouragement. In addition, we would like to thank all the study participants for their perseverance and trust.

Finally, I am deeply grateful to my family for their unwavering support and belief in me, as I stand on the shoulders of their sacrifices and love to become the first in our family to attend college, and proceed to earn a doctorate.
# Contents

Abstract ................................................................. ii
Preface ................................................................. iii

1 Introduction ............................................................. 1
   1.1 Who Gets to Compute? ........................................... 1
   1.2 Let’s Get Creative ............................................... 3
   1.3 Contributions .................................................. 4

2 Data-Driven Circuit Education ....................................... 9
   2.1 Introduction ..................................................... 10
   2.2 Related Work .................................................. 13
      2.2.1 Classroom Management and Learning Tools ........... 13
      2.2.2 Circuit Prototyping Tools ............................... 14
      2.2.3 Teaching Coding Style ................................. 15
   2.3 Formative Study ................................................. 16
      2.3.1 Procedure ................................................ 16
      2.3.2 Participants ............................................. 17
      2.3.3 Results .................................................. 17
      2.3.4 Design Considerations .................................. 19
   2.4 CircuitStyle ..................................................... 20
      2.4.1 Tutorial Authoring Tool ................................ 21
3 Tangible Circuits

3.1 Introduction .................................................. 41
3.2 Related Work .................................................. 44
    3.2.1 Circuit Prototyping and Educational Tools ............... 44
    3.2.2 Tangible Interaction for BVI Persons ................... 45
    3.2.3 STEM Education Tools for Visually Impaired Persons .... 46
3.3 Study 1 ....................................................... 47
    3.3.1 Results ................................................ 47
3.4 Study 2 ....................................................... 48
    3.4.1 Magnitude of Tutorial Accessibility Limitations ........ 49
    3.4.2 Results ................................................ 50
3.5 Design Considerations ........................................ 51
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.6 Tangible Circuits</td>
<td>53</td>
</tr>
<tr>
<td>3.6.1 Interaction Design Overview</td>
<td>54</td>
</tr>
<tr>
<td>3.6.2 Implementation</td>
<td>54</td>
</tr>
<tr>
<td>3.6.3 Automatic Parsing</td>
<td>56</td>
</tr>
<tr>
<td>3.7 Study 3</td>
<td>58</td>
</tr>
<tr>
<td>3.7.1 Participants</td>
<td>58</td>
</tr>
<tr>
<td>3.7.2 Apparatus</td>
<td>59</td>
</tr>
<tr>
<td>3.7.3 Task</td>
<td>59</td>
</tr>
<tr>
<td>3.7.4 Testing session</td>
<td>60</td>
</tr>
<tr>
<td>3.7.5 Procedure</td>
<td>61</td>
</tr>
<tr>
<td>3.7.6 Data Analysis</td>
<td>63</td>
</tr>
<tr>
<td>3.7.7 Findings</td>
<td>63</td>
</tr>
<tr>
<td>3.7.8 Designing Accessible Hardware Computing Tutorials</td>
<td>68</td>
</tr>
<tr>
<td>3.8 Future Work</td>
<td>70</td>
</tr>
<tr>
<td>3.9 Conclusion</td>
<td>70</td>
</tr>
</tbody>
</table>

4 PokerFace

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Introduction</td>
<td>73</td>
</tr>
<tr>
<td>4.2 Related Work</td>
<td>75</td>
</tr>
<tr>
<td>4.2.1 Computationally Augmenting the Body</td>
<td>75</td>
</tr>
<tr>
<td>4.2.2 Masks, PPE, and Facial Cues</td>
<td>76</td>
</tr>
<tr>
<td>4.2.3 Augmenting the Face with Computation</td>
<td>78</td>
</tr>
<tr>
<td>4.2.4 Literature Gap</td>
<td>78</td>
</tr>
<tr>
<td>4.2.5 Implementation</td>
<td>79</td>
</tr>
<tr>
<td>4.3 User Evaluation</td>
<td>80</td>
</tr>
<tr>
<td>4.3.1 Method</td>
<td>80</td>
</tr>
<tr>
<td>4.3.2 Participants and Data Collection</td>
<td>82</td>
</tr>
</tbody>
</table>
7.3.4 Findings ................................................. 182
7.3.5 Design Goals ........................................... 184
7.4 Erato System Design ...................................... 185
  7.4.1 Accessibility Accommodation Requests/Etiquette Guide .... 186
  7.4.2 Speaker Support Features ................................. 189
  7.4.3 Host Exclusive Features ................................. 192
7.5 Validation Study ............................................ 194
  7.5.1 Participants ............................................. 194
  7.5.2 Methodology ............................................. 197
  7.5.3 Task ................................................... 197
7.6 Results ..................................................... 199
  7.6.1 Evaluation of Individual Tools ............................ 201
  7.6.2 Emergent Socio-Technical Behaviors and Dynamics ....... 204
  7.6.3 Design Insights ......................................... 205
7.7 Discussion .................................................. 208
  7.7.1 Anonymous Accommodation Request and Communication ... 208
  7.7.2 Gender Equality and Accessibility Beyond d/DHH .......... 208
  7.7.3 Human to Human Feedback ................................ 209
7.8 Limitations and Future Work ................................. 210
  7.8.1 Mitigating Bad Actors and Potential System Conflicts ...... 210
  7.8.2 Embarrassment and Plurality ............................. 211
  7.8.3 Meeting Session Size and Scope ........................... 211
  7.8.4 Extension to ASL Signers ................................ 212
  7.8.5 Session Host Tools ....................................... 212
  7.8.6 Personalizing Feedback for Discussions .................... 213
  7.8.7 Communal Agglomeration of Accessibility Etiquette Taxonomy 213
List of Tables

2.1 Compiled list of circuit styles aggregated from instructor interviews with average importance score associated with each style. . . . . . . . 18

3.1 Details of circuits used during error identification task. . . . . . . . . 62

4.1 Descriptive statistics of performance by condition including Confederate (C), Participant (P) and Combined Confederate and Participant (CP) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 84

6.1 Participant-suggested use cases from formative Pantomime Study with accompanying operations and interactions. Interaction refers to the input method used by the participant during the pantomime and Operation refers to the behavior performed by the interaction. . . . . . . 138

6.2 Occurrences of specific operations presented by participants during their use case in Phase 2 of the study. Full list of operations can be found in Appendix B . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 139
7.1 Participant table for members selected from survey and interview participants for the mock-classroom formative study. Participant noted with a * symbol indicates that this person served as the session host for the study. Fields noted as N/A indicate participant declined to answer this survey question. Full table of participant characteristics included in Survey and Interview portions of the study can be found in Appendix E.

7.2 Participant table for members selected for the Erato prototype case study where P Number indicates anonymized participant number. Participant noted with a * symbol indicates that this person served as the session host for at least one session of the study.

7.3 Summary of accommodation requests made during the Erato case study.

E.1 Description of participants in formative studies. Fields noted as N/A indicate participant declined to answer this survey question. An abbreviated version of this table containing only participants included in the live study was described in Chapter 7.

F.1 Benefits and pain points of Microsoft Teams as described by DHH/HoH survey and interview participants in Formative Study (See Chapter 7).

F.2 Additional benefits and pain points of Zoom compared to Teams and Meet as described by DHH/HoH survey and interview participants in Formative Study (See Chapter 7).

F.3 Benefits and pain points of Google Meet as described by DHH/HoH survey and interview participants in Formative Study (See Chapter 7).
List of Figures

2.1 Overview of the CircuitStyle interface. (A) Students sort steps of the tutorial; (B) Style authoring tool; (C) Instructor interface for at-a-glance awareness of classroom performance; (D) Student in-situ tutorial interface with dismissible style guide. .................................................. 9

2.2 (A) Example of good circuit style: wires go around ICs instead of over, components lay flat against board, etc. (B) Example of poor circuit style: wires long and tangled, components cramped together, etc. . . 11

2.3 Overview of tutorial authoring; (A) Instructors can preview their uploaded circuit diagram; (B) Multiple photos can be uploaded to create a slide show; (C) Instructor can input additional instructions and step details. ................................................................. 21

2.4 Style Authoring Tool; (A) New styles can be created or default styles can be modified; (B) Styles verified using either quiz or peer review; (C) Description of proper style implementation; (D) Style details; (E) Photo examples of both good and poor style implementation can be uploaded. ................................................................. 23
2.5 Style card from left of student tutorial screen in Figure 2.1D; (A) Expand/collapse style guide button; (B) Overview of style information and proper implementation; (C) Background color corresponded with severity level; (D) Photograph examples of both good and poor style execution.

2.6 The Peer Review Interface; (A) Name of particular circuit style to be evaluated; (B) Photograph example of good style implementation; (C) Photograph submitted by fellow student to be reviewed; (D) Student feedback options.

2.7 Collapsible step hierarchy from Instructor’s in-situ tutorial screen in Figure 2.1C. Steps are color coded according to density of students currently completing that step.

2.8 Detail of intervention methods from Instructors in-situ tutorial screen in Figure 2.1C; (A) Student information and latest photo submitted as part of peer review; (B) Pie chart reflecting student’s successful performance of circuit styles; (C) Intervention methods (e.g., thumbs-up, direct messaging).

2.9 Likert scale responses for CircuitStyle case study. Onscreen tutorial: Overall, students were enthusiastic about following along with the instructor’s pre-written tutorial. Many participants expressed that this was their favorite aspect of the system. We also noticed that having both written instruction and a picture was helpful. Some participants reported relying more heavily on the image than the text, and vice versa. One student initially ignored the tutorial and style guide altogether because they were confident in their abilities, but struggled later to complete the circuit and eventually found several missed steps.
3.1 Overview of the TangibleCircuits system. A) 3D model parsed from Fritzing Diagram; B) Fritzing Diagram used as input by the system; C) User interacting with the 3D printed model’s audio and tangible feedback. ................................................. 41

3.2 Overall results of magnitude accessibility assessment for 3910 web tutorials. Values indicate number of tutorials containing specified media. 49

3.3 A user interacting with the TangibleCircuits prototype. ............... 53

3.4 Audio interface of TangibleCircuits displayed on a commodity smartphone. ................................................. 54

3.5 TangibleCircuits audio interface displayed on a commodity smartphone overlaid with the 3D printed circuit model. ......................... 56

3.6 Sample Arduino Project Hub tutorial. A) Component list; B) Descriptions of components inside the component list. ......................... 59

3.7 Circuits used during circuit error identification task with 2 errors each. A) Simple web stage circuit; B) Complex web stage circuit; C) Simple TangibleCircuit stage circuit; D) Complex TangibleCircuit stage circuit. .................................................. 64

3.8 Average Success Rate of Simple (S) and Complex (C) Circuit and Wiring Error Identification Task. ............................ 65

3.9 Averaged Likert between 1 and 5 with 1 meaning ‘not at all’ and 5 meaning ‘very much’. .................................................. 67

4.1 Four User-Study Conditions: a. traditional mask; b. PokerFace mask with video but without captions; c. PokerFace mask with video and captions; d. PokerFace mask with captions but without video; ....... 74

4.2 PokerFace Prototype Sketches with Dimensions: a. front piece of mask with slot for smartphone; b. back of mask with endoscopic camera ... 80
4.3 Results of user study comparing confederate and participants scores, reflecting the same trend across both datasets. ........................................... 83

4.4 Results of Likert Scale questions from exit questionnaire and interviews. 86

4.5 Averaged results by trial reflecting comparable performance by participants and confederate over time. ......................................................... 87

5.1 Top-down view of objects generated in the branching paths of the Calliope system. ........................................................... 95

5.2 Conceptual Illustration of the Calliope system featuring a collection of objects generated during design walkthrough ....................... 96

5.3 Interaction Overview of the Calliope System: A. An empty vitrine; B. Object generation menu; C. Filled vitrines with generated objects; D: Editing menu; E: Mesh editing; F: Progress lights signal mesh manipulation interpretation; G: Computation complete; H: Results of mesh manipulation interpolation displaying in preceding room. ........... 102

5.4 The Calliope Virtual Environment Design Was Informed By Radial Data Visualization. ................................................................. 106

5.5 Micro vs Macro Creation Viewing; A. User views a single design; B. User lowers walls to view all designs from this session. ............... 110

5.6 Overview of the Calliope User Interface: A. Minimap with favorited object highlighted; B. Vitrine prior to object generation; C. Vitrine containing generated object ......................................................... 111

5.7 Color-Coded System Indicator communicates transparent system load status. ................................................................. 112

5.8 Results of Design Session Indicating Breadth and Depth of User Creations. 113

5.9 Example Objects Generated Using Calliope: A. Custom chair; B. Car Body; C. Sofa Desk; D. Artwork ........................................ 114
6.1 An overview of CLIO’s presentation capabilities for real-time performances, presentations, and storytelling. (a) The presenter creating behavior mappings between input interaction methods (right) and output behaviors (left). (b) A menu containing images and other virtual objects. (c) The presenter selects an image using a mid-air gesture and drags it across the screen. (d) To make an image larger, the presenter uses a two-handed pan/zoom gesture. (e) The presenter uses voice commands and mid-air gestures to display text labels on the screen and draw notes on the image. .................................................. 127

6.2 CLIO’s rehearsal-to-performance workflow derived from formative study.142

6.3 CLIO’s behavior mapping system. (a) Presenter is able to activate mapping menu at anytime during rehearsal or presentation. (b-1) Inside the menu, the presenter selects one operation from the left side of the menu (b-2) the presenter selects one interaction method from the right side of the screen. .................................................. 144

6.4 State-based dynamic gesture detection system. (a) Finger-joint angles used for hand pose detection. (b-1) We detect one of eighteen static hand poses. (b-2) We then look for sequences of static hand poses to detect the beginning and ending of dynamic gestures. (c-1) First state gesture in the sequence (closed fist). (c-2) Second state gesture in the sequence (extended index finger). (d) User performs dynamic gesture. (e) User ends the dynamic gesture by returning their hand from the second state pose to the first state pose. ........................................ 145
6.5 Operations Supported by **Clio**. (a1) Arrange: the user places their hand over the image and creates the Arrange gesture. (a2) the user moves the image to the desired location and stops making the Arrange gesture to place the image in the desired area. (b1) Pan/Zoom: the user places both hands over the image and makes the Pan/Zoom gesture. (b2) the user moves their hands away from each other to enlarge the image. (c1) Draw: the user makes the draw gesture by raising their index finger. (c2) The user draws on the screen with their index finger. (d) Display Text: the user makes the Display Text gesture (raises two fingers) and speaks to display their spoken words as text on screen. (e1) Conjure: The user makes the Conjure gesture (e2) Images appear on screen where Conjure gesture was initiated. (f1) Dismiss: The user makes a dismiss gesture. (f2) Image disappears when user makes Dismiss gesture over visible image. (g) Highlight: similar to draw gesture, the user makes the highlight gesture (three fingers) to draw highlights on the screen text. (h1) Next/Previous: the user makes the Next or Previous gesture (thumb right and thumb left respectively) to hide all presently visible media and text and move to the “next” or “previous” image in the media library, similar to moving to the “next” or “previous” slide in a slide presentation.

6.6 Average Likert Results from Validation Study for Presenters (PR) and Audiences (A) on a 5 point scale where 5 indicates “very much so” and 1 indicates “not at all”. Axis labels reflect components of **Clio**. Color and size of bar indicates average Likert score reported for each dimension evaluated (See Legend).
7.1 Overview of Erato Accessibility Accommodation Tools: (A) Participants and hosts submit anonymous requests in the text input. (B) Submitted requests are listed for the host to approve or reject. (C) Buttons allowing hosts to approve or reject submitted requests. (D) Approved requests are listed for all participants. (E) Participants and hosts can report a violation of the request, the number next to the bell indicates the count of violations reported by participants during the session. (F) Recently approved requests and reported etiquette violations will appear as separate notifications in the Notification Center.

7.2 Additional Participants’ and Hosts’ Features (A) Participants and hosts can monitor their spoken volume by glancing at the textual feedback and speech volume meter. Numerical measurement of speech volume also displayed for transparency. (B) Buttons to provide direct feedback to speaker regarding spoken volume. (C) Playback participant’s or host’s voice into headphones. (D) Push a one-time message into all participants and hosts’ notification center. (E) Participants and hosts can monitor their speech rate by glancing at the textual feedback. Numerical measurement of speech speed also displayed for transparency. (F) Buttons to provide direct feedback to speaker regarding speech speed. (G) Notifications are displayed in the notification center.
7.3 Host Exclusive Features: (A1) Toggle switch to discourage participants from using the text chat to communicate. Once turned on, an automated text notification will appear in the chat display if Erato detects that participants are using the text chat. (B) Participants and hosts are reminded every 40 seconds to turn on their closed captions unless Erato detects that the participant or host has captions turned on. (C) Hosts can monitor the amount of time each person has spoken during the session, as well as see visual indicators in green of who has spoken most recently. (D) Hosts can see the ratio of communication occurring via voice or text usage in this pie chart.

7.4 5 Point Likert Scale Response From Participants Who Are d/DHH to the Question “The Prototype Tool Made Me Feel More Confident While Presenting”. Results Reported are on a Scale of 1 to 5 with 1 meaning “not at all” and 5 meaning “very much so”.

7.5 Average and standard deviations for Systems Usability Scale Results Reported by Participants in our case study. Results are on a scale of 1 to 5 with 1 meaning “not at all” and 5 meaning “very much so”.

A.1 Average computational idle time for single tasks.

A.2 Average computational idle time for parallel tasks.
Chapter 1

INTRODUCTION

Advancements in computing technology such as AR/VR/MR/XR, multi-device, computer vision, and generative systems have led to a paradigm shift in human-computer interaction, with a growing focus on multimodal computing. I will call this paradigm shift towards “omnimodal” systems, and expand the concept below. A ‘mode’ is a type of communication, and humans receive communication using their senses. Multimodal systems, therefore, often revolve around 5 categories of sensory communication: Visual, auditory (audio), haptic (direct/physical), kinetic (motion-based), and ambient (indirect). We can think of a multimodal interaction as an exchange between two parties (a device and a human being) where multiple input or output modalities may be used simultaneously or sequentially depending upon context and preference.

Section 1.1

WHO GETS TO COMPUTE?

The way we build machines determines how they are used and who is able to use them. Multimodal computing combines various forms of interaction, such as voice, touch, gesture, and visual cues, to create more inclusive, efficient, and contextually intelligent technological solutions. This diversity in interaction methods
caters to different preferences and abilities, creating a more natural and seamless user experience. This thesis explores the space between and across modalities, and explores how multimodal systems enrich the expressiveness of computing and broaden the accessibility of our machines. Multimodal systems are defined by their flexibility. Without the ability to transition between inputs and outputs, all a multimodal system provides is a single choice up front about an engagement. However, the true potential of multimodal systems is realized when they transcend mere flexibility, evolving into omnimodal systems that explore the interstices between modalities as fertile ground for innovation. The transition between different modalities, often referred to as an “interaction cliff,” poses significant challenges in maintaining the cohesion and fluidity of the user experience. Whether these transitions are voluntary, based on user preference, or involuntary, dictated by system requirements, they are pivotal moments that can make or break the effectiveness of a multimodal system. These systems move beyond the sum of their individual modes and create new design spaces between their input methods. They point towards a speculative future of machines capable of input and output, in effect, any-modal or omnimodal.

Omnimodality refers to an interaction framework where a person can employ any of their senses in combination to interface with a system. The focus is on creating computing environments that are adaptable to any mode of interaction preferred by the user, recognizing the diversity of human sensory and cognitive capabilities and seeking to design systems that are inclusive and accessible to all. By leveraging the full spectrum of sensory modalities, omnimodal systems aim to provide a seamless and intuitive interface between humans and technology, where the mode of interaction can dynamically adjust to suit the user’s needs and context.
Section 1.2

LET’S GET CREATIVE

The work in this thesis is concerned with creativity, a term historically fraught with ambiguous and contested definitions. Prior work has explored a divergent, social, and mutable definitions of creativity [224]. This invokes familiar notions like the divergent/convergent loop described in design theory, suggesting that creativity comes from diverging to explore as many solutions as possible, before converging on usable solutions [179]. Another common definition of creativity is “rearrangement of pre-existing elements into something novel that reveals something fundamental about the world or ourselves” [134]. Neurobiological, behavioral, and systems neuroscience offer approaches to the structural and neural basis of creativity. Defending a concrete definition of creativity is beyond the scope of this thesis, yet a working definition is useful for such an abstract concept. For our purposes and the purposes of this thesis, we will refer to “creativity” simply as “the ability to make things”.

Do I claim that omnimodal systems can solve all design problems? Absolutely not. In fact, many parts of my research show that developing these systems can often be overwhelming and intimidating to learn, often limiting, not supporting, innovative thinking. It is for this reason that to enable the systems described in this work to support and expand a human’s “ability to make things”, it is imperative to follow the principle of “low threshold, high ceiling, and wide walls” [197]. That is, the tools should be approachable, but should allow their users to make anything they want, and be generic enough to enable a variety of target applications.

I want to empower tools that enable people to express themselves creatively and to develop as creative thinkers. The aim here is to develop advanced multimodal systems and user interfaces that not only boost users’ productivity but also foster
innovation. This objective encompasses a wide range of potential users, including software engineers, various scientists, product and graphic designers, architects, educators, students, and more. Enhanced multimodal interfaces could facilitate more efficient searching of intellectual resources, foster better collaboration among teams, and accelerate the discovery process. These sophisticated interfaces should also offer robust support for hypothesis formation, faster evaluation of alternatives, enhanced understanding through visualization, and more effective dissemination of results. For creative endeavors that involve the creation of novel artifacts such as computer programs, scientific papers, engineering diagrams, symphonies, or artwork, these enhanced multimodal interfaces could aid in exploring solutions, avoiding unproductive decisions, and enabling straightforward revisions.

Section 1.3

CONTRIBUTIONS

This thesis describes 6 systems exploring the tensions between modalities in many-modal systems. Using these systems, blind and visually impaired students learn about engineering and prototyping creative coding electronics [52], designers use their hands to sculpt with generative algorithms like clay [238], and video conferencing participants with disabilities can collaborate and communicate effectively using body driven augmentations [48] [51]. The goal of this research in the near term is to describe what intelligent systems are possible and desirable, and over the long-term drive work in machine learning and computer science in directions to address high-impact social problems. The works included in this thesis include;

(a) Chapter 2 explores how multiple modes of data can be used to support cooperative classroom learning and encourage peer-to-peer mentoring on challenging material [49]. We explore how educational computing can leverage vi-
sual, auditory, tangible, and ambient (social) modes of interaction to facilitate electronics learning and support the development of best practices. A prototype system, CircuitStyle, probes how multiple modes of input and output can enrich and support the needs of instructors and students, creating a collaborative space for learning and making.

(b) Chapter 3 describes an algorithm called TangibleCircuits which automatically translate 2D circuit diagrams into 3D models and a corresponding audio interface [52]. These models could then be 3D printed and attached to a commodity smartphone with the audio interface uploaded to the display. Touching the components of the device would relay audio information regarding circuit connectivity and implementation, resulting in an accessible multimodal circuit diagram. To ensure that the 3D model could be parsed through touch, I deployed additional empirical studies to understand how tangible information differed from visual representation. I developed a model of how geometric, structural, and spatial information could be extracted from tangible stimulus and how this differed across participants. Insights from this model were used to inform and modulate the parameters of my algorithm to redistribute the layout of circuit components and their tangible properties, optimizing for tangible legibility. We learned that multimodal tools generated by the algorithm not only democratized STEM education for blind and visually impaired students but were strongly preferred by non-blind students as well.

(c) Chapter 4 presents PokerFace, an innovative mixed-reality mask designed to improve communication by leveraging multimodal communication during the COVID-19 pandemic, where traditional masks became a daily necessity. By using a smartphone and recycled materials, PokerFace displays a live-stream of the user’s mouth and nose on the mask surface, maintaining safety while en-
1.3 Contributions

Enhancing visibility and clarity of speech using multimodal augmentations to the mask frontpiece. A user study with 18 participants, involving a charades-like game, showed that the PokerFace prototype facilitated better communication compared to traditional masks. The study highlights the potential of embedding computation into face-worn wearables to support interaction and improve communication, especially for those who are deaf or hard of hearing. The prototype explores the design space of mixed-reality masks and face-worn wearables, suggesting applications such as live translation, expression through XR filters, and assistance for individuals with facial paralysis. The research indicates that mixed-reality masks like PokerFace could provide practical solutions for communication challenges in various social situations, even beyond the pandemic context.

(d) Chapter 5 examines how we can engage the senses and our bodies as sources of interaction to democratize access to powerful emerging technologies. Until recently, generative algorithms remained inaccessible to most people because they were difficult to train and required substantial technical knowledge to implement and effectively practice. Even with the proliferation of text-based interfaces for Transformers and other generative technologies, users still have to filter through thousands of candidate solutions generated by the system and continually refine their query in a cross-modal (text to image) approach.

To move past the “command line” era of generative AI, this research explores multimodal systems that enable users to interact, fine-tune, and refine queries to generative systems using whatever modalities best suit their preferences and abilities. We propose a virtual reality suite of tools to facilitate rapid 3D modeling and ideation called Calliope [238]. Calliope enables designers to use traditional sculpting techniques such as carving, pinching, and extrusion as input
to a 3D generative AI. We also developed a suite of tools that take advantage of the 3D nature of VR and the vector arithmetic capabilities of generative methods. Informed by formative studies with designers, artists, and novice users, we adapted existing 3D GAN and VAE architectures for use in VR, as well as developed novel conditional sampling techniques required for the interactions elicited from our target end-users. By embodying the experience of human-AI creation in VR, we found that users were able to design desirable 3D models better and faster than traditional tools or text-based approaches alone. In addition, users built trust with the system and were open to exploring unexpected ideas produced by the generative algorithm. This line of research aims to democratize access to advanced generative AI tools and demonstrates how multimodality can make models more interpretable and steerable for users.

(e) Chapter 6 demonstrates how multimodality can also enrich our communication and accessibility. CLIO is a system that augments video conferencing with a blend of computer vision and natural language processing. CLIO enhances interactions by integrating real-time video and audio augmentations and allows users to control the interface through a combination of voice, body movements, and external devices like smart watches and tablets. This multimodal approach was found to support a diverse range of users, including those with motor or visual impairments, speakers who are DHH, and users of American Sign Language (ASL). The insights from Erato were pivotal in shaping CLIO’s design, leading to a system that could adapt dynamically to users’ evolving needs and preferences. CLIO observed user’s modality preferences and adapted the interface to anticipate how different modes shaped user needs.

(f) Finally, in Chapter 7, we expand upon this research to explore how multimodal systems can greatly enrich communication for people with disabilities. We ex-
amine the limitations of current video conferencing software for people with a broad range of disabilities, with a particular emphasis on the Deaf and Hard of Hearing (DHH) community. Our empirical study included a detailed analysis of commercial platforms, assessing both physical and financial accessibility. This led to the development of a prototype suite of tools, called Erato, designed to overcome the identified accessibility barriers. Remarkably, our evaluation studies revealed that these tools had applicability beyond the DHH community, addressing broader social communication issues. For instance, one participant noted the value of our system in providing objective data on gender dynamics in conversations, highlighting its potential in promoting more equitable and inclusive communication practices.

It is the hope that exploring these systems evidences the need for further investigation and development of computing that engages our senses in new and imaginative ways. Developing modalities between and across the senses opens new possibilities for different bodies and abilities while simultaneously supporting creative problem solving and the development of innovative thinking. It is the hope that this research can use the tools of human-centered design to describe what intelligent systems are possible and desirable, and over the long-term drive work in machine learning and computer science in directions to address high-impact social problems and support innovative problem solving. It is the conclusion of this thesis, that the path towards this paradigm is inherently made of many modes. The future is omnimodal.
Instructors of hardware computing face many challenges including maintaining awareness of student progress, allocating their time adequately between lecturing and helping individual students, and keeping students engaged even while debugging problems. Based on formative interviews with 5 electronics instructors, we found that many circuit style behaviors could help novice users prevent or efficiently debug common problems. Drawing inspiration from the software engineering practice of coding style, these circuit style behaviors consist of best-practices and guidelines for implementing circuit prototypes that do not interfere with the functionality of the circuit, but help a circuit be more readable, less error-prone, and easier to debug. To examine if these circuit style behaviors could be peripherally enforced, aid an in-
person instructor’s ability to facilitate a workshop, and not monopolize instructor’s attention, we developed CircuitStyle, a teaching aid for in-person hardware computing workshops. To evaluate the effectiveness of our tool, we deployed our system in an in-person maker-space workshop. The instructor appreciated CircuitStyle’s ability to provide a broad understanding of the workshop’s progress and the potential for our system to help instructors of various backgrounds better engage and understand the needs of their classroom.

Section 2.1

INTRODUCTION

In breadboard circuit prototyping, circuit style (akin to coding style in programming) refers to a set of rules that uniforms the appearance and construction process of a breadboard circuit to make it readable, understandable, and maintainable (Figure 2.2). Examples of good circuit style behaviors include avoiding crossed wires while prototyping, using as little wire as possible, and checking the polarity of components before insertion. Practicing good circuit style behaviors results in breadboard circuits that are less error-prone, easier to debug, and easier to share. Traditionally, breadboard circuit style has only been taught in universities or colleges to students pursuing a degree in electronics or related fields. However, increasingly, many novice and untrained users, such as in maker communities, are experimenting with breadboard prototyping on their own to incorporate electronics into art projects. In addition, the formal education backgrounds of high school electronics teachers, maker instructors, and workshop facilitators has also broadened beyond the traditional background of electronic engineering [21]. Unfortunately, the majority of tutorials and teaching materials available to these new learners and instructors consists of lessons on traditional electronics prototyping and focuses less enforcing appropriate circuit styles
2.1 Introduction Data-Driven Circuit Education

Figure 2.2: (A) Example of good circuit style: wires go around ICs instead of over, components lay flat against board, etc. (B) Example of poor circuit style: wires long and tangled, components crammed together, etc.

[161]. Given the increasing diversity and evolving needs of both educators and novice users in the hardware computing community, there is need for lightweight learning tools that support users from non-traditional backgrounds and promote useful habits for electronic prototyping.

Using a user-centered approach, we first conducted semi-structured interviews with 5 instructors from various backgrounds to understand the current practice of teaching breadboard circuit style and any challenges the instructors face. From these insights, we assessed the importance of circuit styles and constructed a compiled list of common physical computing best behaviors (Table 2.1), and weighed the importance of the individual behaviors within this list of circuit styles. Although the instructors agreed that enforcing circuit style behaviors was an important aspect of an electronics education, most instructors did not have the time nor the ability to encourage and
2.1 Introduction

Data-Driven Circuit Education

reinforce these behaviors for student individually. To address the challenges delineated by the instructors, we created CircuitStyle, a workshop management tool to help instructors construct hardware prototyping tutorials for in-person workshops, keep track of participant behavior, and peripherally reinforce good circuit style behaviors without monopolizing the instructor or participant’s attention (Figure 2.1). Unlike other workshop or classroom management tools, this paper focuses on the peripheral applicability of circuit styles to follow-along in-person workshop tutorials. We provide a list of circuit styles extracted from literature and interviews with workshop instructors. In addition, our system supports several features which peripherally reinforce these styles for students. These tools aim to encourage classroom engagement and offload style reinforcement to the software and student interaction. For our evaluation, we deployed our system to a makerspace workshop-like environment and conducted a field study to evaluate the effectiveness of our system. We found that our system helped the instructor better engage with their students by reducing the amount of attention monopolized by tracking student progress and reminding students of common behaviors. The instructor also appreciated the tutorial authoring tool and its ability to deepen the instructors understanding of the course material and better anticipate common errors their students may encounter. In addition, workshop participants agreed that the tutorial system was helpful for navigating the process of circuit implementation and receiving circuit style reminders was helpful for debugging their circuit. The main contributions of this chapter’s work are:

• insights into the current challenges of teaching breadboard circuit style in makerspace workshops;

• a classroom management system to help workshop instructors peripherally enforce circuit style behaviors without interfering with participants’ learning of functional circuit construction or monopolizing instructor’s attention; and
2.2 RELATED WORK

- insights from a case study investigating users’ initial impressions of the system’s usability and usefulness.

We discuss several insights for future research in HCI to better support style behaviors for electronics prototyping.

Section 2.2

RELATED WORK

This work builds upon existing research in the design of classroom management and learning tools, circuit prototyping tools, and insights from teaching coding styles.

2.2.1. Classroom Management and Learning Tools

Several commercial products already exist which give an instructor access to the activity of their students’ screens (e.g., Softlink and NetSupport School). These systems make the instructor aware of each student’s activity on the computer and provide the instructor with coarse intervention options, such as freezing a student’s input, or taking control over their computer. Another class of tools in research also provides general support for coordination in the classroom. For example, GroupScribbles [146] extends the concept of sticky notes to digital classroom media. FireFlies2 supports cognitive offloading through the use of tangible pixel devices distributed through the classroom [242]. In contrast to these tools, our system provides contextual information about students’ activity in a specific hardware skill being taught. The idea is to help instructors with early detection of potential problems, and to develop a series of robust “best-practices” to prevent errors. Closely relevant to our project is Maestro designed for in-person 3D modeling tutorials [84] [86] where the tutor sees a dashboard displaying each learner’s editor and can assess their progress. Our system takes inspiration from Maestro’s approach, but examines a different and largely
2.2 Related Work

unexplored domain: hardware prototyping. This alters the problem in several key ways (1) monitoring student progress is difficult because it requires knowledge of a physical object being used by the student, and not a virtual environment or system, (2) the system is meant to be used as a support device that peripherally re-enforces good practices instead of providing the principle means of learning the material, and (3) support tools should ambiently aid the instructor, and students, not monopolize their attention. Finally, a number of software learning systems have used data from software logs to enhance software tutorials [124] [193], or provide improved help or capabilities within feature-rich software [37] [73] [84]. These projects, however, focus on an individual learning or using the software on their own. We are interested in exploring software systems that support learning for hardware computing in group settings, a topic that has only received limited attention [61] [137].

2.2.2. Circuit Prototyping Tools

Prior work has shown that novice users face substantial difficulty in designing and building physical computing systems [21] [161]. Some challenges include choosing correct components, wiring components together, programming logic, specifying variable nomenclature, and debugging. Several research systems have been developed to address these challenges. For example, Toastboard [54] is an intelligent breadboard that assists novices with debugging through LED indicators on the board itself, and a software interface that provides troubleshooting tips. Bifröst [158] instruments both the hardware and software components of embedded computing projects to help users trace the system state and assists in debugging. Trigger-Action Circuits [5] enables users to specify desired functionality at a behavioral level, and generates designs and corresponding instructions for assembling them. PICL [74] allow users to create sensor-based interactive systems using “programming by demonstration” (i.e. using demos to view actions and modifying them for use). Other systems teach fundamen-
tal concepts of circuit design, and programming. For example, Programmable Bricks [196] allows children to develop electronic hardware using LEGO bricks embedded with computers, sensors, and actuators. ElectroTutor approaches this problem by integrating interactivity into traditional step-by-step tutorials for hardware prototyping on the Arduino [251]. Finally, a number of systems have been developed that aid in sensing the state of the electronics components in embedded systems [54] [223], data which could aid in debugging and troubleshooting.

Unlike the systems focused on developing novel hardware and sensing techniques or improving individual instruction, our work examines how circuit style practices can be communicated to novices and reinforced peripherally. We use the metaphor of software coding style as inspiration for developing our system. We also examine this problem from the instructor’s perspective and our approach supplements the in-person mentorship provided by an instructor.

2.2.3. TEACHING CODING STYLE
Evidence in the literature suggests that effectively teaching and enforcing coding style in programming (e.g., indentation, whitespace, naming conventions, etc.) significantly mitigates the number of bugs in a programmer’s code, preemptively prevent programmers from making common errors, and promotes the readability of the code [6]. Although coding style has little effect on the program’s behavior, it does have a significant influence on sustainability and readability for developers [22] [182]. The choice of coding style is largely a matter of developer preference and evolves from their programming experience. Although compliance with coding standards across an institution or project team can enhance team communication, reduce program errors, and improve overall code quality, developers and students do not consistently follow such conventions [139]. Another class of tools assist in enforcing good coding styles. For example, Foobaz is a tool that allows educators to provide custom feed-
back to students on variable names at scale [78]. Similarly, PeerStudio allows students to receive feedback from fellow students, reducing wait-time for help and improved learning [119]. Style Avatar visualizes student’s source code style as facial expressions to peripherally reinforce programming concepts [141]. AutoStyle is a research system that provides automated, adaptive style hints which suggest syntax shortcuts and code skeletons that enforce good coding style [40]. The above systems show that enforcing various aspects of good coding style can improve readability, portability, and maintainability of code while reducing the rate of error. This early experience of reading quality code and experiencing less frustration while debugging is especially crucial for novice users [139]. To better understand if this style behavior could be equally useful within the domain of physical computing, we considered the design space of how style suggestions could be integrated peripherally into circuit prototype training in a group setting.

Based on the above literature review, our first goal was to understand instructors’ current use of hardware computing style protocols when teaching novice users.

2.3.1. Procedure

We devised a semi-structured interview protocol and recruited 5 instructors who taught electronics prototyping at various institutions including formal primary education classrooms, makerspace workshops, and higher education. We examined common teaching tools used, difficulties instructors faced when facilitating in-person circuit tutorials, and common style behaviors they repeatedly reinforced to their students. We
also presented the instructors with a list of potential stylistic choices for hardware computing and asked the instructors to rank their importance on a 7-point Likert scale (Table 2.1).

2.3.2. Participants
The instructors in our study had a variety of backgrounds including Visual Art, Physics, Electrical Engineering, and English. Some had very little experience in teaching hardware computing before they were asked to begin conducting tutorials on circuit prototyping. This was very surprising to us, so we asked the instructors to elaborate further how they learned the material they taught in their workshops or classes. They reported that often they learned the material while preparing for their lecture, often completing the circuit themselves the night before class. In this way, some of the instructors learned some of the material in tandem with their students.

2.3.3. Results
Most of the instructors agreed that circuit style behaviors were important to learn, but difficult to teach because they required repeated reinforcement. Most of the instructors, for example, mentioned the need to repeatedly remind students to check the polarity of various components before inserting the component into their breadboard, or to use a multimeter to ensure their component is working properly. The interviewed instructors agreed that enforcing these behaviors was important, but doing so in a workshop setting required considerable time and attention.

We also found that most instructors had directly taught their students some of the stylistic protocols we identified through our literature review, even though they had not previously been aware of the concept of “style choices” for circuit prototyping. For example, almost all of the instructors repeatedly reminded students to complete the implementation of their circuits before powering their system. Instructors varied
Table 2.1: Compiled list of circuit styles aggregated from instructor interviews with average importance score associated with each style.

<table>
<thead>
<tr>
<th>Circuit Styles</th>
<th>Severity (1-7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ensure component’s polarity is correct before insertion.</td>
<td>7.00 (σ=0.00)</td>
</tr>
<tr>
<td>Avoid changing components on a breadboard whenever the board is powered.</td>
<td>6.88 (σ=0.22)</td>
</tr>
<tr>
<td>Measure the component’s value (resistance/ capacitance/ inductance) before insertion.</td>
<td>6.63 (σ=0.41)</td>
</tr>
<tr>
<td>Check IC part number before insertion.</td>
<td>6.50 (σ=0.61)</td>
</tr>
<tr>
<td>Build and test in subsections.</td>
<td>6.50 (σ=0.87)</td>
</tr>
<tr>
<td>Begin by placing the ICs first. Then connect relevant components which directly reach out from the IC pins.</td>
<td>6.38 (σ=0.54)</td>
</tr>
<tr>
<td>Always check the whole circuit and connect the power supply to the circuit last</td>
<td>6.25 (σ=1.30)</td>
</tr>
<tr>
<td>Verify the power supply voltages and input signals with an oscilloscope or voltmeter.</td>
<td>6.13 (σ=0.54)</td>
</tr>
<tr>
<td>Use the power rails to connect power supply.</td>
<td>6.13 (σ=0.74)</td>
</tr>
<tr>
<td>Push the component down firmly until it cannot go any further.</td>
<td>5.88 (σ=0.74)</td>
</tr>
<tr>
<td>Avoid laying wires or components over ICs.</td>
<td>5.88 (σ=1.14)</td>
</tr>
<tr>
<td>Use as little wire as possible.</td>
<td>5.63 (σ=0.96)</td>
</tr>
<tr>
<td>Ensure wires/components are trimmed to lay flat against the breadboard.</td>
<td>5.50 (σ=1.06)</td>
</tr>
<tr>
<td>Don’t insert two pins of different components or two wires into the same socket.</td>
<td>5.50 (σ=2.06)</td>
</tr>
<tr>
<td>Keep pin 1 of all IC’s pointing the same direction.</td>
<td>5.38 (σ=1.47)</td>
</tr>
<tr>
<td>If more than one IC is involved, make sure they are separated by several rows.</td>
<td>5.25 (σ=1.03)</td>
</tr>
<tr>
<td>Avoid placing two components with long legs close to each other.</td>
<td>5.00 (σ=1.46)</td>
</tr>
<tr>
<td>Run a circuit simulation before building.</td>
<td>4.50 (σ=1.12)</td>
</tr>
<tr>
<td>Understand the breadboard connection (Watch out for split power rails) before you start.</td>
<td>4.50 (σ=1.50)</td>
</tr>
<tr>
<td>Color code the wires of your circuit (e.g., red for power, black for ground).</td>
<td>4.50 (σ=1.80)</td>
</tr>
<tr>
<td>Carefully check the component’s row and column number before inserting into board.</td>
<td>4.13 (σ=1.75)</td>
</tr>
<tr>
<td>Avoid cramming components into compact areas; use the whole breadboard space uniformly.</td>
<td>3.88 (σ=1.88)</td>
</tr>
<tr>
<td>Keep the relative position of components as similar to the diagram as possible.</td>
<td>3.75 (σ=1.89)</td>
</tr>
<tr>
<td>Avoid crossing wires.</td>
<td>3.38 (σ=1.85)</td>
</tr>
<tr>
<td>Begin by connecting power and ground rails.</td>
<td>3.25 (σ=1.79)</td>
</tr>
<tr>
<td>Bend each wire at 90°.</td>
<td>3.00 (σ=2.03)</td>
</tr>
<tr>
<td>Use software to plan the breadboard circuit first.</td>
<td>2.63 (σ=1.71)</td>
</tr>
</tbody>
</table>
significantly in their severity rankings for the compiled list of circuit style rankings (summarized in Table 2.1). Some argued, for example, that it was extremely important for students to consistently color code their wires, while others insisted this was not necessary.

We also found that instructor’s attention was significantly strained during lectures, and as a result of this, enforcing these stylistic behaviors or other best practices is difficult:

[P4] *It’s frustrating... when you just told the entire class [to check the polarity of their components] and then you immediately get a question from one student about why [their circuit] isn’t working... and then the same questions from another student... only to discover that they didn’t check the polarity of the resistors.*

This frustration reflects the exhausting demands placed upon the instructor to reinforce these practices in addition to facilitating the class. A reinforcement system that assists the instructor in peripherally supporting these style behaviors could potentially alleviate the instructor’s burden while assisting new students in developing good implementation stylistic practices.

### 2.3.4. Design Considerations

Based upon our literature review and our interviews, we synthesized the following criteria for designing a new system that reinforces circuit style practices.

*Glance-able awareness of student’s progress.* Since an instructor’s attention may be monopolized by lecturing and assisting individual students, she should be able to monitor a student’s progress throughout the duration of a workshop in a quickly-digestible manner. An instructor should preemptively identify and prevent common errors students may encounter without compromising with their focus.
2.4 CircuitStyle Data-Driven Circuit Education

*Supplement, not replace, an in-person instructor.* Although log data can be useful in enforcing good style practices, this approach cannot capture the nuanced understanding of individual student skills and motivations like an in-person instructor. A style enforcement tool should allow the instructor to decide which stylistic choices should be enforced, as well as when and how these style choices will be enforced.

*Reinforced Mastery.* Since many instructors do not have formal training or expertise in electronics prototyping and circuit styles, these instructors should be able to reinforce their own knowledge and mastery of styles. This design consideration is intended to allow instructors to find as many mistakes as possible in order to prevent these same mistakes in their student’s work.

*Ambient Nature.* To prevent distraction from the main task, a style reinforcement tool should be displayed ambiently. This can reduce the time-cost and required user interaction to access the help content and not add any cost when not using the guidance system (i.e., it can be easily dismissed). To minimize distraction, users should be rarely interrupted.

To address the above design considerations, we implemented CircuitStyle, a web-based tool to reinforce good circuit styles without monopolizing the student or teacher’s attention. CircuitStyle allows instructors to author circuit tutorials and use the classroom management feature for live instruction, and a student interface for following tutorials and carrying out peer reviews.
2.4 CircuitStyle Data-Driven Circuit Education

Figure 2.3: Overview of tutorial authoring; (A) Instructors can preview their uploaded circuit diagram; (B) Multiple photos can be uploaded to create a slide show; (C) Instructor can input additional instructions and step details.

2.4.1. Tutorial Authoring Tool

CircuitStyle first allows instructors to interactively compose a step-by-step tutorial for students to follow (Figure 2.3). In addition, an instructor can assign circuit style behavior guidelines to individual steps in the tutorial, and decide how these behaviors will be evaluated.

Integration with Existing Tools. The instructors are first asked to indicate which circuit they would like to build. Many of the instructors that we interviewed suggested that their tutorials were often taken from websites such as Sparkfun or generated using the open source Fritzing software. To accommodate this, we allowed instructors to upload a csv file containing the steps of their tutorial generated by Fritzing. This also includes uploading a picture of the completed circuit for student reference.
**Style Authoring Tool and Default Styles.** Our formative study showed that the breadth and importance of certain style choices varied greatly from instructor-to-instructor. To account for this, the second step of our authoring tool allows instructors to create their own style behaviors, or modify a list of default stylistic choices (Figure 2.4). Previous work indicated that coding styles are most effective when consistent across a project or organization, and malleable according to the instructor’s needs [138]. Thus, allowing the instructors to modify the stylistic choices from project to project provides the flexibility instructors desire with the consistency students need.

Each circuit style (Figure 2.5) contains information pertaining to the proper implementation of the style, and indicates the level of severity which can adjust student’s attenuation to various styles [139] [194]. In addition, instructors can upload photographs of good and bad implementations of the style which has been demonstrated to better engage students and assist in students being able to distinguish between good and bad circuit style implementation [139] [217]. Instructors can choose between validating their circuit styles through either a quiz or a student-driven peer review depending on which method they feel better evaluates that particular style [78]. Default style choices were compiled from our formative instructor interviews and aggregated according to their severity rankings. (e.g. wire color coding, connecting ground and power rails, etc.) These default styles are also editable by the instructor according to their needs.

**Authoring and Scalability of Tutorial Complexity.** Next we ask the instructor to organize the steps of their tutorial into subsections since we observed that this was done mostly manually by instructors. This step helps the overall organizational flow of the workshop, especially when working on larger, more complicated circuits. We designed this feature to help streamline this process and help scale workshops to larger circuits.
The instructors then step through the tutorial themselves and assign behaviors to each step. Since instructors may come from a variety of backgrounds, many of them may need to complete the tutorial themselves to fully understand the material. Our system accounts for this by encouraging tutorial authors to photograph their own circuit after each completed step. This is optional and can be replaced by images from Fritzing or other software. We encourage instructors to participate in this aspect of the tutorial authoring system, as mistakes made by the instructor could prove useful in preventing the same mistakes being replicated by students. The purpose of this step is to reinforce the instructor’s understanding of the material and allow the instructor to teach a representational sample of coding styles to better familiarize students with good style habits [139]. In addition, this allows instructors to automate the style reinforcement process by scheduling trigger events that reinforce circuit style behaviors [220]. By completing the circuit themselves instructors may be better aware.
of which stylistic behaviors could prevent common mistakes.

2.4.2. Interactive Live Tutorials for Students

The student interface consists of a web application in tandem with a mobile phone application for live workshops. Our goal was to peripherally reinforce good circuit style behaviors in hardware computing for students without monopolizing too much student or instructor attention.

**Step Sorting.** Once the instructor has completed authoring their tutorial, students can log into the phone application and web interface. Once logged-in, students are presented with an overview of the tutorial (Figure 2.1A). Many of our interviewed instructors expressed frustration with ensuring student engagement and attention when overviewing the circuit. To account for this, students are met with an image of the finished circuit, accompanied by a randomized sorting of the tutorial steps. Students are asked to sort the steps into the correct order before continuing. This peripherally reinforces common wiring procedure steps, as well as ensuring that students pay attention to the instructor as they walk students through the circuit.

**Step-by-Step Tutorial and Style Guide.** Students are next guided through the construction of the circuit step-by-step as authored previously by the instructor (Figure 2.1D). The left expandable panel contains information regarding the particular style behavior indicated by the instructor during tutorial authoring for that step (Figure 2.5). This includes a detailed description of the style, and photos contrasting good and bad examples of the style behavior. Having access to style references has been documented in the literature as an effective and commonly-used method for incorporating style behaviors into programming since it mitigates the amount of memorization demanded by the student [217]. For this reason, we have included this
2.4 CircuitStyle Data-Driven Circuit Education

feature here to facilitate a similar reference experience for hardware computing.

**Peer Review.** To mitigate the amount of attention demanded by the instructor when evaluating these stylistic behaviors, we designed a peer evaluation system to allow students to review each other’s work and provide feedback (Figure 2.6). This experience is akin to the coding style review process in the software engineering industry, and has demonstrated effectiveness at engaging students to learn through example and practice [139].

At the end of each section, students are asked to photograph their circuit. We chose to conduct the peer review process during section breaks to mitigate the participant’s cognitive load that may result from code switching. Students are then presented with photos of other student’s work, and asked if a particular style is evident in the photograph. Students can choose between “yes”, “no”, and “I don’t know” in addition to leaving a comment (Figure 2.6). Encouraging peer-to-peer and peer-to-instructor interaction has been shown to be an effective reinforcement technique for coding styles [138].

The number of circuits a student is asked to review at the end of each section depends on how many students are in the class, and how many styles the instructor designated for examination during tutorial authoring. This aims to reinforce the student’s understanding of the peripheral style materials, mitigate potential errors, and reduce the amount of attention needed by the instructor. In addition, this practice is intended to encourage students to identify good circuit style behaviors, which has been demonstrated in the coding style literature to be an essential skill for developing good style behaviors [217]. After the student’s peer review is completed by their fellow students, a Style Reminder slides-in from the top right hand corner indicating which styles the student performed well and possible violations.
Figure 2.5: Style card from left of student tutorial screen in Figure 2.1D; (A) Expand/collapse style guide button; (B) Overview of style information and proper implementation; (C) Background color corresponded with severity level; (D) Photograph examples of both good and poor style execution.

**Style Reminder.** After the peer review phase, students will receive feedback in the top right corner slide-out window which describes which styles they successfully performed and which they may have missed. A collapsible style guide directs the students to review any styles they may have missed, or performed poorly during their last peer review. The style reminder feature works in tandem with the collapsible style guide and the peer review system to reinforce good style.

### 2.4.3. Student Progress Observational Interface

During live workshops and tutorials, we provide an interface for instructors to gauge workshop behavior and progress at-a-glance (Figure 2.1C). Our goal in the design
of this system was to support the instructor by providing essential information on student performance, while not monopolizing instructor attention. This consists of 3 principal components described below.

![Peer Review Interface](image)

**Figure 2.6: The Peer Review Interface; (A) Name of particular circuit style to be evaluated; (B) Photograph example of good style implementation; (C) Photograph submitted by fellow student to be reviewed; (D) Student feedback options.**

**Tutorial Progress.** The right hand panel contains an expandable hierarchy of the steps and sections in the tutorial as authored previously by the instructor (Figure 2.7). Next to each step is a number indicating how many students are currently completing that step of the tutorial.

In addition, tutorial steps and sections are color coded according to the following scheme: grey indicates that no students are currently working on this step, red indicates that very few (≤ 0.25) are currently working on this step, and green indicates
that the majority of students are currently working on this step \( (\geq 0.50) \). This color scheme was designed to allow instructors to momentarily glance at the screen, and have an understanding of the class’s overall progress through the tutorial [44].

![Collapsible step hierarchy](image)

Figure 2.7: Collapsible step hierarchy from Instructor’s in-situ tutorial screen in Figure 2.1C. Steps are color coded according to density of students currently completing that step.

**Student Submissions.** The center of the screen can be button-toggled to either reflect information regarding a specific step, or view student’s most recent submissions for peer review (Figure 2.1C). The tutorial view allows the instructor to view the information pertaining to any particular tutorial step, including instructions and photographs related. By pressing the toggle button, the student submission window displays the most recently submitted student photos for peer review, as well as the student’s name, and a pie chart reflecting their successful performance of circuit style behaviors as evaluated by peer-reviews and quizzes (Figure 2.8A).

This is intended to provide awareness of each students past behavior history,
current status [44]. The screen also allows instructors to provide positive feedback
to students in the form of a thumbs-up, direct intervention in the form of freezing
the student’s screen, constructive feedback via comments, as well as mitigate various
issues regarding the peer review process such as contested reviews (Figure 2.8C).
We designed this feature to encourage individualized feedback that complements the
automated triggered feedback of quizzes and peer review [119] [139] [220].

Behavior Performance. The bottom panel contains a series of pie charts (Bottom
of Figure 2.1C) showing student stylistic behavioral performance (the green indicates
the proportion of students who successfully completed that stylistic behavior during
the previous peer-review). These charts are sorted from worst-performing behavior
to best-performing behavior, allowing the instructor to gauge which behaviors might
need to be further reviewed. Providing awareness of overall classroom performance
is necessary to understanding areas of needed reinforcement and direct intervention
[139].
Section 2.5

USING CIRCUITSTYLE IN PRACTICE: A CASE STUDY

We deployed CircuitStyle in a workshop setting to investigate whether circuit style behaviors could be peripherally enforced by our system and to gather users’ initial impressions of the system’s usability and usefulness.

2.5.1. Procedure

We recruited 11 students for a 90-minute instructor-led electronic prototyping workshop where students were asked to create a circuit and used CircuitStyle for assistance. The instructor was asked to construct a typical makerspace circuit tutorial using our interactive authoring tool. We asked the instructor to narrate their experience using our tool as they proceeded with the tutorial composition process using our software. The resulting circuit tutorial guided students through the creation of a button-activated, battery powered motor circuit. The instructor was given a brief overview of our software before construction their tutorial.

Students were then brought into the workshop area and asked to complete a brief preliminary questionnaire regarding their previous experience with electronics and awareness of coding style practices, as well as circuit style behaviors. Next, the students were instructed to log into our web application and phone application and guided by the instructor and software to complete the tutorial. We also recorded video of student progress which helped us extract various quantitative measurements of student performance.

Upon completion of their circuit, students were asked to complete an exit questionnaire and interview. We also conducted an exit interview with the instructor to
better understand their experience with our system.

2.5.2. Study Participants

Our workshop was led by an experienced maker-space workshop leader who also authored the tutorial and guided the workshop. He was recruited from another university due to his extensive experience and interests in conducting mentorship practices in makerspaces and running workshops in hardware computing. His formal education consisted of film studies although he now is pursuing a doctorate in computer science and business. The 11 workshop participants (2 females, 8 males, 1 non-binary) ranged in age from 20 – 29 and possessed some familiarity with electronics. But most participants (63%) did not have a formal background in electronics, even if their current work involved prototyping with electronics. The majority of participants (45%) were studying software engineering or computer science related disciplines while others were concentrating in electrical engineering (27%), music (19%), or design (9%). While almost all (91%) of our participants were familiar with programming coding styles, only 36% were familiar with circuit styles. Those that were familiar, indicated that they had learned these behaviors from instructors or by learning on their own. None of the participants were involved with this project’s research in any capacity beyond participating in the study.

2.5.3. Key Findings from the Workshop

We found that students overall enjoyed the workshop and that the instructor found the tool useful for planning their lesson as well as surveying the performance of the workshop. We discuss our key results by first focusing on the facilitator impressions followed by student feedback.
2.5.4. Facilitator Impressions

In this section, we revisit our design goals presented earlier in context of feedback provided by the workshop instructor.

*Glance-able awareness of student’s progress.* The instructor overall appreciated CircuitStyle’s assistance in keeping track of the overall progress of the class. In particular, he emphasized the utility of the automatic indication of which students had completed which steps. He highlighted that the visual nature of many UI elements helped mitigate the amount of attention he had to attenuate to the system. For example, he found the color coding of individual steps to be intuitive and helpful: *It was useful to see where people were in the steps... the color coding was helpful because it let me know where they were getting stuck without having to look too hard.* The instructor reported that usually with step-by-step tutorials, he had little information on how students were progressing in the project. He was enthusiastic about the follow-along detection feature in CircuitStyle that allowed him to see the most recent photo of the student’s circuit: *I’m a visual person so I liked seeing the individual circuits... this was the most helpful aspect for me because I could immediately see how the students were doing... was really helpful to keep track of where they might get stuck.* In a traditional workshop environment, the instructor would spend a lot of time walking around to see how students were performing. With CircuitStyle, he could spend more time actually helping the struggling students: *[With CircuitStyle]... I could hone in on a few students and help them as opposed to walking around to find out who was struggling.* He also expressed appreciation at our positive re-enforcement tools, such as the ability to give a student’s most recent submission a “thumbs-up” or send a quick note. He indicated that novice students often are unsure if they completed a step correctly, and this allowed him to provide positive re-enforcement of good work.
**Supplement, not replace, an in-person instructor’s capabilities.** In terms of supplementing in-person instructor capabilities, we were surprised to find that the instructor expressed an increase in their engagement with the class. In interacting with our behavior performance feature, he mentioned that although it was a secondary task to examine the pie charts at the bottom of the screen, it was helpful to see if one particular style was severely being missed: *I liked that it showed me potential challenges students might face...because they weren’t following the style guide...and seeing individual student circuits told me who needed help.* Despite the instructor’s enthusiasm for our system, he indicated that the system might be even more helpful for students in larger workshops. In our workshop, we noticed that if a student encountered a problem, they would usually ask their immediate neighbor for assistance in debugging the problem and not always rely on the peer review. Examining the usefulness of this feature in learning contexts where one-to-one interaction may be more difficult for students and instructors is a key area for future investigations.

**Reinforced Mastery.** We asked the instructor to comment on the interactive authoring tool and its ability to help clarify the instructor’s understanding of the material. Overall, he found the step-by-step nature of the authoring tool helpful and appreciated the ease of finding potential pitfalls that students might encounter, and how adhering to a specific style may prevent those errors: *I noticed that students might not wire their ground and power on the same rail...which was good, because connecting the rails with a wire is a good thing to do.* The instructor also mentioned how helpful it was to lay out and design a specific tutorial and learn from their own mistakes in implementing the project before the student: *It was nice because I had time to prepare and it was much more defined...less things can go wrong this way and if things do go wrong, I can spend less time debugging it.* He also suggested that adding features to allow note-taking at various levels might be helpful as well. Both
while prepping the tutorial and while facilitating the workshop, the instructor mentioned that they took copious notes which became unwieldy to organize according to step, section, and overall workshop activity.

**Ambient Nature.** A key concern of this project was to not monopolize or further strain the instructor’s attention with our system. The instructor expressed appreciation of the peer-review system’s ability to enforce these behaviors without relying on persistent instructor intervention: *I didn’t have to worry about identifying who was struggling... was good because I knew who needed help from my screen and could go straight to them.* In addition, the instructor expressed that the overall UI reduced the burden of enforcing good circuit implementation practices by peripherally encouraging these behaviors with our system.

### 2.5.5. Student Feedback
All students were able to successfully complete the circuit well within the allotted 90 minutes. The mean completion time of the total circuit tutorial was \( \mu = 41\text{min} 16\text{s} \) (9min 14s SD). Only 3/11 students who used the style guide and followed the step-by-step tutorial system required additional assistance from the TAs outside of the instructor’s lecture. Questionnaire feedback indicated that students were overall enthusiastic about the workshop, rating their enjoyment an average of \( \mu = 4.1 \) out of 5. Figure 2.9 shows a summary of students’ responses to individual components of the interface on a standard 5 point Likert scale based on the following criteria, i) Helpful, ii) Distracting, iii) Confusing, iv) Difficult, and v) Engaging.

**Sorting step.** We asked students if the sorting step was helpful for reinforcing their understanding of proper procedure. Overall, our results indicate that many students (in particular, those with little to no background in electronics) found this step to be
particularly confusing (See Figure 2.9). Our main motivation for creating the sort-ordering task was to make it easier for instructors to tell their students to complete their circuits in a proper order (e.g., inserting ICs first and powering the circuit last). Although somewhat useful, our results showed that novice students found this sort-ordered task to be confusing.

![Average Likert Scale Results of Case Study](image)

Figure 2.9: Likert scale responses for CircuitStyle case study. Onscreen tutorial: Overall, students were enthusiastic about following along with the instructor’s pre-written tutorial. Many participants expressed that this was their favorite aspect of the system. We also noticed that having both written instruction and a picture was helpful. Some participants reported relying more heavily on the image than the text, and vice versa. One student initially ignored the tutorial and style guide altogether because they were confident in their abilities, but struggled later to complete the circuit and eventually found several missed steps.

**Onscreen style guide.** Participants were enthusiastic about the onscreen style guide, noting that it helped keep their circuit tidy and organized. Most students, even those with electrical engineering training, referenced the style guide while constructing their circuit and found it helpful: [P8] *It was good because most of these things I*
learned by making mistakes...some of [the styles] I learned before and it was a nice review, but some I had not learned yet. Many novice users also indicated that the style guide helped them feel more secure in their circuit and reassured that they were progressing through the tutorial adequately.

**Performing Peer Review.** In an attempt to minimize the attention needed by the instructor to reinforce style behaviors, we required the students to perform peer review. Students overall reported that performing peer review helped reinforce their understanding of the material, but wondered if it was useful in such a small class: [P4]... it was useful to see how others were doing... but I think this might be more useful in a larger class... it might also be difficult with more complicated circuits since you won’t really be able to see [the style] from a picture. Novice users also reported feeling insecure about the helpfulness of their comments to other students. While participants with formal training in hardware computing found providing feedback to be an intuitive process, novice users were unsure if their input was helpful: [P2] I wasn’t sure if my comments were helpful since I didn’t have much experience working with circuits... I might say something wrong. Although students have an option to select “I don’t know” and are not obligated to leave a comment, students may feel obligated to provide more input than required.

**Receiving Peer Review Feedback.** Participants overall appreciated receiving a peer reviewed report, but felt that the current system did not provide enough feedback. Our current method of delivering feedback to students involves a small scroll-in window in the top right-hand corner of the screen to make it as unobstructed as possible. However, many students reported looking for further information on their performance and not being able to locate it.
Receiving Circuit Style Reminders. We also asked participating students to evaluate our style reminder system and found that this feature helped students identify potential problem areas while debugging: [P5] *I had trouble getting my circuit to work at the end but remember that one [style reminder] said that I hadn’t connected my ground rails... it was great. I knew where to debug and get it working.*

Section 2.6 DISCUSSION

We have contributed the design of CircuitStyle, a tool for peripherally reinforcing circuit style behaviors for in-person workshop tutorials. We have also demonstrated the value of techniques which assist instructors in facilitating workshop activity, as well as laid the preliminary work for exploring the applicability of circuit style behaviors to improving hardware computing education. Although we only examined CircuitStyle’s utility in a single small workshop setting, a more in-depth investigation into how this tool impacts instructor and students’ experience at a large scale can provide additional useful insights. In this section, we discuss some limitations and avenues for future work.

2.6.1. Scaling Peer-Review Features

When asked to complete peer review, several students asked their neighbors to evaluate if the work had been done correctly. This was unexpected, and a practice that could be encouraged by a differently configured peer review system. In future work, it would also be useful to understand the performance of our current system at a larger scale workshop. Similarly, our peer review system may prove difficult to scale for more complicated circuits since identifying style behaviors from a single photograph of a complex circuit may be challenging, especially for new users. This issue could
also possibly be alleviated by encouraging in-person peer evaluations as opposed to virtual peer evaluations.

2.6.2. Accounting for Varying Skill Levels

Some students also expressed insecurity in their ability to provide useful feedback to their fellow students, particularly if the student was a complete novice user. We could account for this by calibrating our system prior to the workshop with user background information, and establish a “virtual mentoring” system by encouraging more experienced users to provide feedback to less experienced users. Finally, our system requires the instructor to author a step-by-step tutorial for the students, and thus our current system is only reliable for follow-along workshops. Evaluating our system’s usability in free-form workshop environments is a principle area of future work.

2.6.3. Alternative Workshop Structures

Although our system proved sufficient for pre-structured tutorials, not all workshops employ follow-along instruction methods for teaching hardware computing. Since instructor and student activity in these free-form workshops differs significantly, additional design considerations must be accounted for employing our system in this domain. Additionally, deploying our system in alternative workshop settings such as formal education classrooms could provide insight into the versatility and longevity of our system. Identifying, adapting, and evaluating our system for such workshop structures is a key area for further investigation.

2.6.4. AR-based Approaches

Incorporating additional input modalities and interaction techniques could further mitigate some of the attention demands of the system. Incorporating AR into our
system could provide additional methods of communicating and peripherally reinforcing circuit style and tutorial material to students, as well as further aiding the instructor in facilitating the activity of the workshop. Additionally, this interaction modality could encourage physical activity and peer interaction during the workshop which could be particularly useful during peer review.

Section 2.7

CONCLUSION

This work provided initial validation for the applicability of circuit styles in follow-along tutorials as well as supported the notion that circuit style behaviors could be peripherally reinforced. Our prototype system and case study evaluated a series of techniques that aide instructors in authoring tutorials, facilitating workshop activity, maintaining awareness of class progress, and reinforcing good circuit prototyping practices without monopolizing instructor attention. Our work lays the foundation for architecting a future where instructors collaboratively share the experience of teaching with a trusted system, allowing the instructor to fully focus on enjoying the mentoring of their students. More broadly, our work calls for more HCI research in the domain of hardware computing to better support the growing number of novice and untrained users.
We present a novel haptic and audio feedback device that allows blind and visually impaired (BVI) users to understand circuit diagrams. TangibleCircuits allows users to interact with a 3D printed tangible model of a circuit which provides audio tutorial directions while being touched. Our system comprises an automated parsing algorithm which extracts 3D printable models as well as an audio interfaces from a Fritzing diagram. To better understand the requirements of designing technology to assist BVI users in learning hardware computing, we conducted a series of formative inquiries into the accessibility limitations of current circuit tutorial technologies. In addition, we derived insights and design considerations gleaned from conducting a formal comparative user study to understand the effectiveness of TangibleCircuits as a tutorial system. We found that BVI users were better able to understand the geometric, spatial and structural circuit information using TangibleCircuits, as well as enjoyed learning with our tool.
In the maker community, novices learn circuits with breadboards by following examples in tutorials from the web. However, most of the existing web tutorials are inaccessible to the blind or visually impaired (BVI) community because they rely heavily on visual information to communicate the material (Figure 3.1B). This is a significant loss considering that members of the BVI community have traditionally been inventors of life-changing electronic devices that benefit both blind (e.g., Optacon) and sighted people (e.g., cruise control) [101]. The high bar of entry to learning electronics excludes the BVI community from participating in innovation via making. BVI children also miss-out on critical STEM education and further high-tech careers [33]. While many accessibility tools exist, most do not encourage or enable BVI users to create their own accessibility tools. In the words of one expert we interviewed for this work:
3.1 Introduction

Tangible Circuits

E1: “Blind people are born makers because the world was not made for them. They have to recreate the world for themselves to thrive.”

Thus, it is our vision that these tools must be designed to support learning for BVI users, enabling and unleashing their creative potential. In this paper, we propose an interactive 3D printed tutorial system, TangibleCircuits, that combines a cost-effective tactile model of a breadboard circuit with audio-feedback for BVI makers and students. TangibleCircuits comprise an automatic parsing tool which translates a circuit diagram (Fritzing format) into a 3D model that is printable with a commercial 3D printer and Proto-pasta Composite Conductive PLA material. The tactile circuit model has components printed using conductive filament and can be affixed to a smartphone to allow for touch-based interaction for learning. When each component or wire is touched, audio feedback details the name of the component, the position, and other details regarding its connection and implementation. TangibleCircuits is intended to broaden the inclusivity and accessibility of maker spaces and engineering classrooms by allowing instructors to create cheap, portable, and easy to use multimodal circuit tutorials. Our vision for TangibleCircuits is to allow tutorial authors and instructors to generate a tangible model and audio interface from existing Fritzing diagrams. These resulting tools can then be 3D printed using a commodity 3D printer and affixed to touch-screen devices to serve as multimodal accessibility tools for BVI students.

TangibleCircuits was developed with a user-centered universal design approach, where a series of studies were conducted to understand the problem space and its magnitude. To begin our investigation, we conducted a semi-structured interview with 3 BVI makers in order to understand several major accessibility issues they encountered using electronics education tools. Examples include the difficulty in understanding the spatial (component layout), structural (debugging), and geometry
information (i.e., component size and shape) of breadboard circuits. In addition, we evaluated the magnitude of these issues by surveying 3910 online tutorials from the most popular open-source tutorial platforms (Arduino Projects Hub and Fritzing Hub). Online tutorials were examined due to their common use as teaching material for novice engineers and makers. We found that that over 98% of online tutorials were not adequately accessible to BVI users according to the Web Content Accessibility Guidelines (WCAG). From these preliminary investigations, we extracted a series of design guidelines for TangibleCircuits. To evaluate the effectiveness of our approach, we conducted a user study with 8 self-reported blind and 6 visually impaired/low vision/legally blind participants, where we evaluated the accessibility of TangibleCircuits and web tutorials modified to be BVI accessible according to WCAG. We found that our system was better at assisting BVI users at recognizing the geometric information, spatial and structural information of components within the circuit. Participants also discussed that TangibleCircuits was fun to use, and significantly less strenuous and frustrating to interact with than online web tutorials.

The main contributions of this work are:

- an understanding of the accessibility issues in the existing circuit learning tools for BVI users;
- an approach to address the issues using interactive tactile models for circuit tutorials;
- insights from a user study, evaluating the accessibility of our prototype and web-tutorials modified to meet standards of WCAG web accessibility.
This work builds upon many intersecting bodies of work including Circuit Protoyping and Educational Tools, Tangible Interactions for Visually Impaired Persons, and Insights from existing STEM education tools.

### 3.2.1. Circuit Prototyping and Educational Tools

Prior work has shown that novice users face substantial difficulty in designing and building physical computing systems [21] [161]. Some challenges include choosing correct components (geometric information), wiring components together (spatial information), and debugging (structural information). Several research systems have been developed to address these challenges. For example, Toastboard [54] is an intelligent breadboard that assists novices with debugging through LED indicators on the board itself, and a software interface that provides troubleshooting tips. Other systems teach fundamental concepts of circuit design, and programming. For example, Programmable Bricks [196] allows children to develop electronic hardware using LEGO bricks embedded with computers, sensors, and actuators. Finally, a number of systems have been developed that aid in sensing the state of the electronics components in embedded systems [5] [49] [158] [223] [251], data which could aid in debugging and troubleshooting.

Unlike the systems focused on developing novel hardware and sensing techniques, our work examines how insights from these techniques can be adapted to enable visually impaired persons to learn electronic prototyping. It is our intention to create a platform that simultaneously employs a universal design approach, as well as ensures the user can learn as autonomously as possible. The purpose of a universal design approach is to similarly enable visually impaired and traditionally sighted users alike
3.2 Related Work

Tangible Circuits

with a single prototype design in order to encourage the tool’s wide adoption. Additionally, our goal of ensuring autonomy is to support the pseudo-autodidactic nature of online learning platforms. For these purposes, tangible and audio feedback systems present a viable modality to achieve these goals.

3.2.2. Tangible Interaction for BVI Persons

Most technologies that are accessible to BVI people substitute visual information with audio-feedback or touch-feedback. Touch is a promising modality for sensory substitution, as previous studies have revealed superior tactile acuity for blind people over sighted people [33]. However, few tangible user interfaces (TUIs) for visually impaired people have been designed, and the existing accessible TUIs mainly broaden accessibility to geographic maps and diagrams. Examples of tangible diagrams include a prototype for the non-visual exploration of graphs and maps by McGookin et al. [157]. These tangibles systems provide multimodal feedback for the creation and modification of diagrams and maps. Other multi-sensory projects include MapSense and IllumiWear [29] [46] which integrated scents (e.g., olive oil, honey) or sound, thus creating a multi-sensory map. More closely related to our interests are tangible maps, where map elements are represented by physical objects which are often augmented with audio feedback [58]. In some cases, users can not only explore the maps, but build and modify them by manipulating and moving the objects. Similarly, the prototype by Schneider and Strothotte [204] enabled visually impaired people to construct an itinerary using building blocks of various lengths with the help of audio cues. Tangible Reels [59] are physical icons on a multi-touch table representing points of interests. The system guided the user with audio instructions to correctly place, link, and retrieve the names of objects.

TUIs have shown many advantages over standard mouse and keyboard computer interfaces. They foster collaboration and have also proved to increase engagement of
students in learning tasks [72] [210]. Moreover, constructing tangible maps improves the understanding and memorization of spatial information in the absence of vision [59]. Similar to our interests is Interactiles which uses conductive 3D printing to increase smartphone devices accessibility [264]. Some preliminary work has been conducted translating these tools into the domain of STEM education tools.

3.2.3. STEM Education Tools for Visually Impaired Persons

Designs for learning computer programing and electronic engineering for BVI users are limited [149] [200] [221]. The few developments in this area include accessible programing languages (i.e. Quorum) and speech interfaces (i.e. Emacspeakiv) that can be effective tools for those who already know how to code, but are less suitable for novices. To assist BVI computer science majors to learn how to program, Smith et al. [218] introduced JavaSpeak, an editor providing additional information about the structure and semantics of written Java code. Other examples include systems which simplify programming logic and provide audio feedback, and tools which help children using screen readers create chatbots [19] [200]. Additionally, Kane and Bigham [105] described teaching BVI students how to analyze Twitter data, producing 3D printed visualizations that allowed for a tactile exploration of their program output. These approaches mostly serve to increase the accessibility of text-based programming by simplifying coding syntax or teaching the use of screen-reader or magnification software. As such, they are more suitable for textual information than visual information. In addition, with engagements being primarily bound to a computer screen, they rarely support hands-on physical engagements. Thus, they do not capitalize on the possibilities offered by manipulating physical objects for learning complicated concepts [213], or for supporting collaborative learning [96].

Despite tangible programming languages and tools for sighted users [60] [96] [159] [230], little work has been done to explore the effectiveness of these modalities within
the realm of physical computing. Some early work in this field is evident in the work of Li et al. who used tactile templates combined with audio feedback to aid users in understanding and manipulating the spatial information of a web-page layout [136]. In addition, some early work documents the potential usefulness of 3D printed models as learning tools for BVI users [213] [212] [204]. Tangible Circuits builds upon these insights in order to design a tangible and audio system for educating novice BVI makers.

Section 3.3

STUDY 1: SEMI-STRUCTURED INTERVIEWS

To aid in our understanding of current practices and needs for accessible circuit prototyping education for BVI engineers, we conducted a semi-structured interview with 3 BVI participants familiar with circuit prototyping technologies. This included a blind engineer who facilitates workshops for BVI people to learn about electrical engineering, a blind technology administrator at a local school for the blind, and a BVI student whom had previously studied physical computing at the college level.

3.3.1. Results

We first wanted to understand current practices in hardware education and found that web-tutorials were often relied upon by our interviewed instructors. Literature corroborated this insight, revealing that web tutorials were commonly used by educators of a variety of backgrounds as a principle source of classroom material [49]. This indicated to us a need to better understand the current accessibility of open-source tutorial systems (see Section 3.4). One of our experts expressed frustration at using these tutorial systems within the classroom. They revealed that upon matriculat-
ing into university, their intention was to pursue engineering as a major, but found that while some accessibility tools made programming easier, navigating circuit implementation was impossible due to the cognitive load required to understand circuit diagrams using a screen reader. This indicated to us the need for multimodal feedback as a necessary design consideration (see Section 3.5). We also inquired about current tangible methods used within this domain, and found that tactile diagrams were commonly used, but due to the abstraction used in direct graphic translation, these diagrams remain largely unusable. Thus, it is imperative for our design to support recognizability of components more suitable for a tactile domain. Finally, we discovered that existing circuit education tools for producing tactile assets included braille embossers and swell paper which are prohibitively expensive and not commonly available in engineering educational settings. This indicated a need to make our tool as inexpensive and ubiquitous as possible, and usable to makerspace and electronic classroom educators with more ubiquitous tools.

Section 3.4

STUDY 2: EXISTING TUTORIAL ACCESSABILITY

As indicated by our interviewees during our semi-structured interviews, web-tutorials often serve as a primary source for classroom material for novice engineers. In order to assess the current accessibility of open-source tutorial systems, we conducted a formative study of web-based hardware computing tutorial resources. The focus of this initial study was to understand the magnitude of accessibility limitations within online open-source tutorial platforms, as well as insights into common web accessibility pitfalls.
3.4 Study 2

Figure 3.2: Overall results of magnitude accessibility assessment for 3910 web tutorials. Values indicate number of tutorials containing specified media.

3.4.1. Magnitude of Tutorial Accessibility Limitations

For the purposes of this study, we collected 7321 online tutorials from two popular online open-source tutorial platforms: Arduino Project Hub and Fritzing Hub. After filtering, this collection comprises 3109 tutorials collected from Arduino Project Hub and 801 tutorials taken from Fritzing Hub. We filtered out tutorials that were empty, not in English, or were significantly incomplete (i.e. missing project description). These tutorials were then analyzed using the Web Content Accessibility Guidelines (WCAG) an online protocol and guideline system for ensuring web accessibility. These guidelines are divided into 4 areas of concentration: perceivable, operable, understandable, and robust. From these guidelines, we extracted 4 characteristics of accessibility which are applicable to hardware computing tutorials. We then assessed the accessibility of the collected tutorials based on this criteria. Results are detailed Figure 3.2.
3.4.2. Results

While 79.5% of the 3910 entries from Arduino Hub contained graphics or photographs, only 2% contained graphic descriptions. This violates the WCAG guideline of Perceivable, weakening the accessibility for users with visual impairments. Furthermore, the preliminary data shows that 53.8% of these tutorials use a circuit diagram and 15.5% use a schematic, but only 3% contain circuit descriptions (see Figure 3.2). An understanding of these visual medias is imperative to completing the tutorials because the circuit diagram (i.e. Fritzing Diagram) communicates spatial (component layout), and geometric (component size and shape) information of breadboard circuits. Both circuit diagrams and schematics communicate structural (wiring) information which is largely missing from these tutorial systems. Only 13.3% contain written step-by-step instructions, and less than 1% contain video with captions. Overall, we found that less than 2% of tutorials surveyed met the criteria for accessibility according to WCAG, indicating a significantly limited accessibility in these online tutorial platforms. Of the tutorials surveyed, we found that the 801 tutorials extracted from Fritzing Hub were less accessible than those from Arduino Project Hub. Fritzing Hub tutorials because they relied heavily on circuit diagrams (Fritzing diagrams) as their primary tutorial material, and lacked textual descriptions of the circuit or components. In fact, 98.2% of Fritzing Hub tutorials contained a circuit diagram, but less than 1% contained circuit or component descriptions. Details from this analysis can be found in Figure 3.2. Furthermore, through this process, we identified key pitfalls of frustration when navigating these media using screen readers. Component descriptions, for example, if included in the tutorial, were usually contained within large HTML tables which were frustrating to navigate using a screen reader. This was largely due to the tables containing information pertaining to the operation of the webpage (such as table indices and tag information) that was not relevant to tutorial material. In
addition, relevant information such as component names was also inaccessible because component names were often extracted from the file name of images associated with the component. This resulted in verbose, unreadable component names that were difficult to associate with a given component.

The results of our study suggest that a system designed to meet this accessibility gap must mitigate the significant difficulty, time, and labor necessary to communicate component descriptions and circuit connectivity to a novice BVI learner. These results motivated us to provide direct access to component information through 3D replicas (see Support Recognizability) and audio feedback of a touched component (see Section 3.5). Furthermore, given that broadening accessibility to these tutorials is a time-consuming endeavor, it is necessary to automate as much of these tasks as possible in order to create a system that is easy for BVI users to understand the circuit tutorial contents.

Section 3.5

**DESIGN CONSIDERATIONS**

Based upon the insights from the above study, we devised a series of criteria to inform the design of our system. From our collected semi-structured interviews and our preliminary study, we devised the following considerations.

*Support Recognizability*. According to our initial study, one of the key components missing from the tutorials examined in our study is adequate description of components. While most tutorials contained a list of components, none contained adequate visual or tactile component descriptions. Furthermore, we learned from our semi-structured interviews that tactile graphics and maps were often insufficient due to their direct translation of abstract graphics to a tangible medium. We thus chose
3.5 Design Considerations

Tangible Circuits

to explore a direct 3D representation of components for our prototype. Any system designed to meet these needs must therefore account for this discrepancy in current tutorial system technology.

**Multimodal Feedback.** A recurring limitation in current tutorials lies in the lack of non-visual communication methods. This discrepancy not only violates the WCAG Perceivable principle, but also excludes populations unable to interpret visual material. Thus, a system designed to account for this limitation must incorporate multiple forms of feedback and guidance, (e.g. audio, tangible, etc.) in order to increase accessibility.

**Support Understanding of Circuit Structure.** A key to understanding the functionality and implementation of a circuit is understanding the structural and spatial information of the circuit, including connectivity of different components and their interactions [21]. According to our experts, this principle is not present in current hardware computing accessibility technologies. In the words of one of our domain expert interviewees: E2: “*Descriptions of circuit diagrams only get you so far, you really need to see how things are put together to get them to work… otherwise, debugging is near impossible*”. Therefore, our system must account for this knowledge gap, enabling users to understand the layout and interaction of various components.

**Automated Accessibility.** As evidenced by our semi-structured interviews and formative study, considerable time and effort is demanded of tutorial designers to meet standards of accessibility. Therefore, it is necessary to automate a portion of the accessibility limitations evident in these tutorials. While the Fritzing platform enables a wider audience of novice engineers and makers to create and interpret circuit diagrams, our previous study indicates that the current interoperability of
3.6 Tangible Circuits

To account for the above design considerations, we created TangibleCircuits: an audio and tangible circuit tutorial system. TangibleCircuits comprises an automated parsing system which translates a Fritzing diagram from a visual medium to a 3D model and voice annotation. This model can be 3D printed using Proto-pasta CDP12805 Composite Conductive PLA material and affixed to a commodity touch-screen smartphone or tablet for voice output. The resulting interactive tactile diagram allows a user to tangibly understand a circuit using touch-triggered voice-feedback.
3.6 Tangible Circuits

3.6.1. Interaction Design Overview
Since audio and tangible feedback have demonstrated effectiveness for communicating information to BVI users, our design focuses on integrating these two modalities of communication. To interact with the tactile diagram, a user simply touches any component, triggering audio information regarding that component to be read to the user. This allows a user to gain insight into the implementation and composition of the circuit while becoming familiar with the tangible shape of each component.

3.6.2. Implementation
TangibleCircuits takes a Fritzing Diagram as its input, and parses the diagram into a 3D model and touch based audio interface. These two complementary components comprise our tutorial system, and the resulting interactive tactile diagram operates on a commodity capacitive touch-screen device, such as a smartphone or tablet without any modification to the device.

Audio Interface. The audio interface consists of a series of buttons laid-out on the display of the touch-screen device (See Figure 3.4). Each button is associated with a different component present within the circuit diagram. When touched, the device

Figure 3.4: Audio interface of TangibleCircuits displayed on a commodity smartphone.
reads audio information related to the target component associated with the button. This information includes the target component name, relevant neighboring components to which the target component is connected, and implementation instructions such as when the component should be inserted. The system repeats this information until the user releases the button, and only responds to a single touch. The user is notified if they are touching more than one component.

**3D Printed Circuit Model.** The 3D model is extracted from the Fritzing Diagram and renders an approximate replica of the components within the circuit. This is intended to provide a tangible approximation that mimics the tactile qualities of the physical breadboard circuit the model represents. TangibleCircuit’s components and wires are printed using Proto-pasta CDP12805 Composite Conductive PLA hard extrusion filament. This conductive filament is crucial to the operation of the device. The breadboard and case is printed separately with non-conductive PLA filament. For our purposes, and for the purposes of reusability we printed the case and circuit separately. These two elements can easily be printed as a single unit using a two filament dual extruder printer as needed. Although the case and board are printed using non-conductive filament and the components printed with conductive filament, both these elements were printed as a single unit using a multi-material 3D printer. This casing allows the tactile circuit diagram to sit above the capacitive touch screen. Each component in the tactile circuit diagram sits directly above its corresponding audio interface buttons, and thus triggers the voice annotation below each component when touched (See Figure 3.5). The resulting interaction allows for both audio and tangible interaction to inform the user of the circuit’s spatial, structural, and geometric information.
3.6 Tangible Circuits

3.6.3. Automatic Parsing

In order to reduce the labor required to create the necessary audio interface and 3D model for each circuit, our system includes an automatic parsing tool which renders the related 3D model and audio interface for each circuit. Our parsing tool first takes a standard Fritzing Diagram as input, which is then parsed for component id tags. These id tags contain the name of the component, its x and y coordinate position within the diagram layout, the pins of the breadboard in which the component is inserted, and the ids of the components connected to the target component. Wires are also described in a similar way, in that their id contains their pin insertion locations and connected components. Since each component and wire id tag contains a series of x, y coordinate positions as well as pin insertion locations, we are able to determine the relative size of the component as well as its relative position on the breadboard. Our system then identifies a 3D modeled component within our component dictionary, comprising a series of component ids and their corresponding 3D models. These component models were taken from open-source online repositories and collected into our dictionary. Once this has been completed, our tool assigns each 3D model component to a location on a 3D breadboard according to the x and y coordinates.
y coordinate positions associated with the parsed component id. The 3D model is then rendered and output as an stl file for 3D printing.

The corresponding audio interface is parsed in a similar manner, where each component is assigned to a touch button whose size and location are determined by the two x and y coordinate positions of each component id. In addition, each component id tag contains information regarding the insertion pin locations for each component, as well as other components within the circuit which it is connected to. This information is parsed, associated with the corresponding touch button, and read using a speech synthesizer. The resulting system allows for input of a Fritzing Diagram, and output of a 3D model and audio interface. Several challenges were involved in translating Fritzing diagrams to 3D representations appropriate for a TUI.

Crossed wires are often present in Fritzing diagrams, but are problematic when translated to a TUI due to the capacitive nature of our interaction technique. If two wires are crossed, it may be difficult for our audio interface to differentiate between the two wires, resulting in confusing feedback. Our system addresses this by identifying potential crossed wires and locating suitable alternatives that maintain circuit connectivity. Similar issues exist for components that are positioned close together in the Fritzing diagram. This could result in a component being difficult to touch in isolation, again resulting in confusing audio feedback. The TangibleCircuits parsing algorithm locates components which are within 2 pins of each other, and considers the quantum of unoccupied pins surrounding the component. If such space exists, the algorithm redistributes the components with 2 pins in between. This ensures that components were spread-out enough to be recognizable through touch. We identified that 2 pins were sufficient for our purposes through a small pilot study with a BVI student. We also determined the dimensions that smaller components, such as wires and resistors, should be printed for tactile recognizability. As a result of this
inquiry, we adjusted the wire thickness to 1mm (scaled larger) and left the component size the same. Scaling components larger actually resulted in greater confusion for the participant as well as complicated our parsing algorithm’s ability to redistribute components effectively.

Section 3.7

STUDY 3: FORMAL USER STUDY

In order to understand the effectiveness of TangibleCircuits for assisting BVI users at understanding sample breadboard circuits, we conducted a formal user study. The focus of this evaluation is to understand how TangibleCircuits complements and contrasts open-source web tutorials at communicating circuit tutorial implementation. Our study consisted of two sessions: learning and testing, as well as two stages: Tangible Circuits and web tutorials. In the learning session, participants were asked to learn a sample circuit using either TangibleCircuits or web tutorials modified to WCAG accessibility standards. The testing session followed the learning session immediately, in which, participants were asked to complete two tasks: a component identification task and an error identification task.

3.7.1. Participants

14 participants (10 female, 4 male) with varying self-reported visual impairments (8 self-reported blind) and electronics educational backgrounds (11 self-reported “none”) were recruited through online advertising, and assistance from a local organization serving the BVI community. Participants ranged in age between 27 and 67 with a median age of 47. Participants were compensated for their time.
3.7 Study 3

Tangible Circuits

Figure 3.6: Sample Arduino Project Hub tutorial. A) Component list; B) Descriptions of components inside the component list.

3.7.2. Apparatus

Study apparatus included the interactive tactile circuit diagrams running on an Android smartphone. For the web tutorial condition, participants were asked to bring their own laptop equipped with their preferred accessibility tools due to the common practice of highly customizing BVI screen reader and screen magnifier interfaces to suit their needs.

3.7.3. Task

Learning Session. For the TangibleCircuits stage, participants were introduced to our interactive prototype’s functionality and usage. We briefly explained the audio feedback mechanism and demonstrated the general use of the device. The audio feedback for each component contained information regarding the components placement and connectivity to other components on the breadboard. We then asked participants to explore the spatial relationship of components, and geometry of components using the tangible and audio feedback of the device. Once participants felt they had a reasonable understanding of the circuit structure, we proceeded to the testing session.

For the web-tutorial session, participants were asked to navigate to an online web tutorial which had modified to meet WCAG accessibility standards and uploaded
3.7 Study 3

Tangible Circuits

to Project Arduino Hub (Figure 3.6). To fully bring these tutorials to accessibility standards, we referenced the Smith-Kettlewell Technical File (SKTF) for examples of hardware computing tutorials designed specifically for BVI engineers. The SKTF is a commonly used reference manual for circuit descriptions and tutorial descriptions for BVI engineers. Using this document as a resource, a member of our research team with a formal background in computer engineering in collaboration with our experts modified these tutorials to meet WCAG and SKTF accessibility standards. These modifications included adding component descriptions, circuit descriptions, step-by-step written implementation instructions, and written video caption transcriptions. Each tutorial contained a list of components needed to implement the circuit, as well as component descriptions we created according to SKTF. In addition, each tutorial contained a written step-by-step direction list for assembling the circuit. Once participants had opened the tutorial, we asked the participant to read over the tutorial using their screen reader, screen magnifier, or other accessibility devices. After the participant felt they had an understanding of the tutorial content, we proceeded with the testing session.

3.7.4. Testing session

Component Identification. For this task, we presented participants with a bucket of 17 common electronic components (e.g. resistors, LEDS, etc.). The bucket contained only 1 example of each kind of component. We then asked participants to use their stage apparatus (web tutorials or TangibleCircuits) as a guide for identifying components used in the construction of the tutorial circuit. Participants were asked to read the name of a component in the component list, and then pull each physical component out of the bucket one-at-a-time, and state whether or not the component they held was the target component from the tutorial. Participants were not told if their identified component was correct in order to prevent learning-effects between the
two stages of the study. We recorded whether or not their choice was correct for each component as well as the time taken to identify the components. After the tutorial components have been correctly or incorrectly identified, we proceed immediately to the circuit error identification task.

**Circuit Error Identification.** We presented the participant with a completed circuit using physical components on an unpowered breadboard (Table 3.1). Each of these physical circuits were similar to the circuit described in the tutorial apparatus (web tutorial or TangibleCircuits), but contained 2 errors: a wire error and a component error. A wire error involves either a misplacement or missing wire, while a component error comprises a missing or replaced component. Participants were then asked to use the tutorial apparatus as a reference for answering three questions regarding the physical circuit: 1. Is this physical circuit the same circuit described in the tutorial? 2. If this circuit is different, how so? 3. How would you modify this physical circuit to match the circuit described by the tutorial? interview. It should be noted that the error modality was the same across stages. In the case of the simple circuits, the two error modes were removed a component and move a component. On the complex circuits, the two error modalities were replacing and removing a component for both the web tutorial and the TangibleCircuit stage of the study. Once the participant has answered these three questions for the circuit, we proceeded immediately to the next phase of testing.

### 3.7.5. Procedure

Each session was 90 minutes long and documented using audio and video recording. Participants were assigned to group A or group B prior to the study. Group A performed the web tutorial stage first, and group B performed the TangibleCircuits stage first. This counterbalance was done in order to eliminate any potential learning-effects.
that might result from our study design. Prior to the study, participants were given a
brief introduction to the functionality of a breadboard and its role as a tool in circuit prototyping. The session began with a demographic and technology experience questionnaire. Participants were then asked to either asked to complete the web-tutorial stage or the TangibleCircuits stage, depending on their group assignment. Each participant completed both the learning session and the testing session for two different circuit tutorials in both the web-tutorial and the TangibleCircuit stage. Following [143], each stage contained one simple tutorial, and one complex tutorial. These 4 circuit tutorials were the same for all participants, and each participant examined the same 4 circuit tutorials. Each stage began with the learning session of the simple tutorial. After participants had completed the learning session for the simple tutorial, we proceeded immediately to the testing session, followed by the learning and testing session of the complex tutorial. Upon completing the first stage of the study, we introduced participants to the apparatus (web tutorial or TangibleCircuit
3.7 Study 3

tutorial) to be used in the second stage of the study. We then immediately proceeded to the learning and testing session for the second-stage simple circuit, followed by the learning and testing session for the second-stage complex circuit. After completing both stages of the study, participants completed an exit questionnaire and interview.

3.7.6. Data Analysis

For the identification task of both the web tutorial and the TangibleCircuit stage, success rates of component identification were recorded, as well as time taken to identify each of the components. During the circuit error identification task, error identification success rate, and correction rate were recorded in addition to time taken to answer each of the three questions posed during the task. In addition, audio and video were recorded, transcribed, and analyzed to evaluate each participant’s understanding of the circuits composition and functionality. We also collected qualitative feedback, as well as Likert-scale usability evaluations as part of the exit interview.

3.7.7. Findings

Overall, participants performed significantly better on the component identification task and the circuit error identification task with TangibleCircuits. In addition, our qualitative findings reflect that participants enjoyed working with TangibleCircuits more than web-tutorials. In this section we revisit our design considerations and discuss how TangibleCircuits services these criteria within our use-case scenario of classroom and makerspace accessibility tools.

Support Recognizability Results. We concluded from our formative studies that direct 3D representation of components as well as providing direct access to component information through touch could better support recognizability of components than screen-reader aided web tutorials. On average, participants identified 62% of
the circuit components with the TangibleCircuits apparatus versus 34% with the web-tutorials. Furthermore, 3 participants who completed the TangibleCircuits stage first were able to correctly identify the resistor component, but unable to do so when subsequently completing the web-tutorial stage. This indicates that overall, geometric information of the components was better recovered by participants using TangibleCircuits than web-tutorials. Furthermore, participants were able to identify 83% of the wiring and component errors with TangibleCircuits versus 27% with the web-tutorials. This indicates that spatial information of the circuit was better communicated using TangibleCircuits as well. Even when using web tutorials as a guide, participants expressed a preference to walk through the tutorial using the physical circuit, touching each component as they progressed. When asked about this, participants expressed the need for a physical guide to accompany the online tutorial information. “It’s a spatial thing, even though I am able to tell where the components are in the tutorial, I would have no idea if they were in the right spot [on the physical breadboard]” (P1). This indicates the importance of tangible communication in understanding the spatial information of the circuit.
**Multi-Modal Feedback Results.** Our formative studies also indicated a need for multimodal forms of communication to mitigate dependency on purely visual media. “I was surprised how much I was able to understand just by touching... I was shocked that I actually could find the errors” (P2). In addition, participants emphasized that they believed they would be more capable of completing the circuit tutorial using TangibleCircuits. However, participants also cautioned that they may not be able to replicate the circuit using TangibleCircuits due to the inaccessible nature of the breadboard itself. Participants also expressed that web-tutorial’s circuit diagrams and circuit descriptions were not helpful, and that touching the TangibleCircuit prototyped was more helpful at understanding the spatial information of the circuit. “The diagrams were useless because I could not see them. I would never be able to complete the steps on my own” (P9). When asked if the audio feedback or tangible feedback was more useful for understanding the circuit’s spatial information, participants insisted that both were equally useful and necessary. “It was great having audio feedback together with touch because together they help better identify the pieces. I am better with touching things” (P11).

Figure 3.8: Average Success Rate of Simple (S) and Complex (C) Circuit and Wiring Error Identification Task.
Support Understanding of Circuit Structure Results. A key component indicated by our formative studies to circuit education is the understanding of circuit structural information such as connectivity. This information is crucial for identifying circuit errors and debugging, and is often lacking for BVI students due to reliance on visual media to communicate this information. We found that identifying the wiring error and component error were completed with different degrees of success. As we can see in Figure 3.8 these tasks individually were performed more successfully with TangibleCircuits than the web tutorials, indicating that structural information of the circuit was also better communicated using our prototype. In addition, we found that participants with total blindness performed differently than those with low vision. We observed that participants with low vision relied more on the visual diagrams of the web tutorials to understand the tutorial material, versus the textual information. These participants had to view the monitor very closely using a combination of screen magnifiers, contrast adjustment software, and screen readers and reported that using the web-tutorials were strenuous on their eyes. Participants with total blindness used screen readers exclusively for the web tutorial stage and overall performed better using the TangibleCircuits device than their low-vision peers. Furthermore, participants expressed that they would prefer to use TangibleCircuits over web tutorials to learn about circuit prototyping. “The audio is real advantage. I know when I touch something, I’m hearing information about that thing... I would never be able to do that with web [tutorials]” (P8). This immediate access to relevant information based upon touch contributes to participant understanding of circuit structure by mitigating the graphic abstraction common to circuit diagrams.

Automated Accessibility Results. Although the automation and design of our tool is intended to mitigate the labor demanded on instructors, we found that some material was not encapsulated within the parsed Fritzing diagrams. This included use-
Figure 3.9: Averaged Likert between 1 and 5 with 1 meaning ‘not at all’ and 5 meaning ‘very much’.

ful component descriptions, which we had to manually insert into our audio interface. We address this issue in further depth in Limitations and Future Work. Although our results suggest that TangibleCircuits could be useful for BVI engineers to understand spatial and geometric information of the circuit, many users still expressed a need for step-by-step instructions in order to feel confident they could replicate the circuit (Figure 3.9). This reflects that in our current implementation, not all necessary information could be extracted through automation, and thus the original tutorial still served as a useful tool for some users. Thus, we conclude that TangibleCircuits serves as a supplementary accessibility tool, but does not completely replace current tutorial technology. Instead, TangibleCircuits narrows the gap of accessibility for these users.

Additional Participant Feedback. Participants reacted enthusiastically to the TangibleCircuit prototype. The results of our 5 point Likert scale (1 meaning not at
all and 5 meaning very much) exit questionnaire demonstrated that participants found TangibleCircuits to be easier to use, less frustrating, and less confusing than web-tutorials (see Figure 3.9). 5 of the 7 participants with legal blindness claimed that the circuit diagrams were the most useful part of the web tutorial. However, we found that these participants averaged 32% correctness when performing the identification task and 13% correctness when performing the circuit error identification task. We noticed that these participants often strained to use their eyes, and even commented that this practice was painful and obtrusive. This indicates to us that these participants were reluctant to trust a tangible medium because of their default reliance on sight. Finally, participants gave several suggestions for how tangible and audio could be better used together for learning circuit prototyping. We detail these in future work below.

3.7.8. Designing Accessible Hardware Computing Tutorials
Since this work constitutes the first effort to create tangible systems for BVI within the domain of hardware computing, we offer the following design insights for further investigations in this field.

**Design 3D Models for Tactile as Well as Visual Use.** A common pitfall during our study was the misidentification of components which were similar in tactile quality (e.g. wires and resistors). This is largely due to the fact that 3D modeled components are designed for visual, not tangible, usability. Any system that uses 3D modeled parts for communicating circuit information must carefully consider the tangible quality of each component and its distinguishability from other components with similar tactile qualities.

**Work With, Not Against, Current Practices.** Many participants expressed insecurity regarding their ability to replicate a given circuit with step-by-step in-
3.7 Study 3

Tangible Circuits

instructions or TangibleCircuits in isolation. This is due to the constraint of having to count pin holes on a traditional breadboard to check proper component placement. By designing to support participants' understanding of circuit structural information, multimodal feedback fills a knowledge gap within current circuit prototyping practices, without diminishing the value of those practices themselves.

**Cost Effective Solutions Through Tertiary Users.** We found during our formative studies that a key to the adoption of accessibility technologies for STEM education lies in their cost-effectiveness. This is largely due to the lack of resource access faced by many BVI engineers, as well as educators potentially not having access to specialized accessibility tools. By considering tools readily available to potential tertiary users (3D printers in maker spaces, smartphones, etc.) we shift the financial burden of creating accessible education and broaden the inclusivity of classrooms.

**Mitigate Information Overload with Gesture.** During our user study, participants suggested that the touch based interaction could be improved by reconsidering our current gesture. In our current implementation, the device continues to read information regarding the selected component to the participant until the user stops touching that particular component. 3 participants mentioned that this relayed too much information, and a multi-tap gesture might work better. Multi-tap would allow different information to be communicated about the component each time it is touched. Furthermore, multi-touch input techniques could also be helpful for allowing a user to touch 2 components simultaneously and receive information regarding their relationship. These considerations of touch input technique remain a promising avenue for further inquiry into touch and audio hardware computing tutorial systems.
3.8 Future Work

LIMITATIONS AND FUTURE WORK

The promising results of this initial work indicate many avenues for future investigation. Although our parsing tool can automatically construct an audio interface using the information in the Fritzing diagram, these files often do not contain all details necessary for BVI users to understand a given circuit. For this reason, our audio interface required some manual input of missing information including component color and usable component names (e.g. “green wire” vs “wire 5”). However, this problem could be easily mitigated by embedding this information within the id tags of the file itself using techniques such as [186]. Furthermore, TangibleCircuits is suited for small circuits which are not egregiously complicated. The majority of web tutorial circuits are simple, suitable for novices to use for learning fundamentals. It is our vision that more complex circuits could be explored and implemented using TangibleCircuits by decomposing large, complicated circuits into smaller, modular elements which could be integrated to implement the larger system. Future work will explore algorithmic techniques to implement this decomposition process. Finally, in order to ensure the universality of our design, we intend to deploy a similar user study with sighted users.

Section 3.9

CONCLUSION

We present the magnitude of accessibility limitations novice BVI engineers face in understanding the geometric, spatial and structural information within the domain of hardware computing. Through a semi-structured interview with 3 BVI makers as well as formative studies, we compiled 4 design considerations to inform the construction of a multimodal tangible and audio interface for replicating breadboard circuits.
3.9 Conclusion

called TangibleCircuits. This system comprises an automatic parsing algorithm which takes a Fritzing Diagram as input, and renders a 3D model and touch-based audio interface as output. These two elements are combined to create our interactive device which fits a capacitive smart-phone form factor. Our formal user study indicates that TangibleCircuits mitigates the accessibility gap of web-tutorials, and is enjoyable for BVI students to use. We believe BVI users bring valuable perspectives to hardware computing and push for greater inclusion of their voices and insights. It is our vision that BVI engineers will design and construct their own accessibility devices in the future.
The 2020 COVID-19 pandemic made masks a daily wearable for personal protective equipment as a public health precaution. Traditional mask designs obscure communication by obstructing the face and muffling the voice which can make communication especially difficult for users who are deaf or hard of hearing (DHH). PokerFace uses a commodity smartphone and recycled materials to display a live-stream of a user’s mouth and nose on the mask surface. This maintains the safety precautions afforded by the mask, while mitigating the obfuscation of traditional mask designs. To compare PokerFace’s ability to facilitate communication with traditional masks, we conducted a user study with 18 participants, who played a collaborative communication game similar to charades. Participants performed better at this collaborative communication task with our prototype than with traditional masks, and even non-DHH users became aware of the importance of lip-reading and facial cues in communication due to study participation.
During the COVID-19 pandemic response, masks for personal protective equipment (PPE) became a daily wearable as a public health precaution in 2020. As a result, communication and socialization was affected by regulations of social distancing and had to adjust to having the face partially obscured by masks. Besides visually obscuring a large portion of the face, traditional masks also muffle the voice of the wearer by covering their mouth with multiple layers of fabric or protective material. These difficulties are exacerbated for users who are deaf or hard of hearing (DHH) since they may rely on lip reading or facial cues to comprehend communication. Subsequently, many innovations began emerging to ameliorate the communication limitations imposed by daily mask usage. This included focusing mask design on personal expression [184], using variable materials to allow the facepiece to be transparent, or even replacing the entire cloth mask material with hard plastic to mitigate visual obstruction to the mouth and nose area. Innovation to mask design was not strictly relegated to design of the mask itself, but also included embedding masks with computation capabilities [169]. This process of imbuing computational capabilities into daily wearables proceeds from a history of ubiquitous computation including smart fabrics, e-textiles [46], and smart jewelry [130][174]. Thus, we explored the potential for the mask as a face-worn wearable embedded with computation and interactivity, as well as the capability of this design space to mitigate limitations in communication resulting from the mask form-factor.

In this paper, we present PokerFace, an interactive, mixed-reality mask which allows for face-worn interaction and expression using a commodity capacitive touchscreen device and cost-effective recyclable materials (Figure 4.1b, 4.1c, and 4.1d).
4.1 Introduction

PokerFace

Figure 4.1: Four User-Study Conditions: a. traditional mask; b. PokerFace mask with video but without captions; c. PokerFace mask with video and captions; d. PokerFace mask with captions but without video;

The mask situates a smartphone in a cradle over the user’s mouth and nose, similar to a traditional mask. PokerFace uses the display of the device to live-stream a video of the wearer’s mouth and nose in real time, as well as present captions of user’s speech using a real-time speech-to-text function. Our goal with this prototype was to explore the following research questions:

- R1. How can embedding computation into face-worn wearables support interaction?

- R2. Can augmented masks facilitate better communication for users than traditional masks?

To explore our research questions, we iteratively designed a prototype mixed-reality mask, and conducted a counterbalanced user study with 18 participants. During the user study, participants were asked to play a series of games of Guess What, a charades-like game, where users try to communicate concepts to each other using gestures and pantomime. We compared across 4 conditions: traditional masks, PokerFace prototype with video but not captions, PokerFace prototype with video and captions, and PokerFace prototype with captions but without video. Our study
suggests that users are able to perform better playing this collaborative communication game with the PokerFace prototype than with traditional masks. The principal contributions of this work include:

- A prototype mixed-reality mask, called PokerFace, made using iterative design process using a commodity smartphone (Section 3)

- Evaluation of prototype PokerFace mask in a counterbalanced study examining its ability to facilitate communication (Section 4)

- Suggesting prototype applications which explore the design space of mixed-reality masks and face-worn wearables (Section 7)

While the prototype studied in this paper is not itself a practical implementation, it demonstrates the potential of exploring this design space while technologies that could make it viable are still being developed.

Section 4.2 RELATED WORK

We first examine the history of embedding computation in wearables and computationally augmenting the body (Section 2.1), before providing an overview of existing research regarding masks for PPE (Section 2.2). Finally, we examine prior work in augmenting the face with computation (Section 2.3).

4.2.1. Computationally Augmenting the Body

Computationally augmenting the body is a popular interdisciplinary domain of research within the field of human computer interaction because on-body embedded systems provide immediate, convenient access to information and services. Early successes of the ubiquitous computing movement imbued daily wearable garments with
4.2 Related Work

computation such as smartwatches and headphones [219]. Research interest in developing the capabilities of these common wearables continues through innovation in interaction [66], biosensing [166], tangibles [56] and other domains. Computational augmentation of the body also takes a wide variety of form-factors from e-textiles [46], and robotic jewelry [106]. Research into virtual, augmented, and mixed reality (XR) augmentation of the body is limited. Prior approaches incorporate displays [130] or projection [90] onto the body to enable interaction or data visualization [259]. XR work in this domain often blurs boundaries between virtual reality and augmented reality. For example, Kim, Lee and Koh incorporated a display onto the back of a virtual reality headset to provide user’s not wearing the VR headset a view of what the user in VR was seeing [114]. Previous work examined how assistive wearables [104] could support end users with hearing impairments (deaf, deafened, and hard of hearing) [102] [165] [172], cognitive impairments [245], visual impairments (such as: blindness [69] and color-blindness [229]), as well as motor impairments [32]. Wearables also present a promising venue for augmenting, restoring, or replacing human skills, abilities, and senses [79][233][173]. While this prior work is promising, little investigation has examined the potential for mixed-reality imbued masks as an interactive face-worn wearable.

4.2.2. Masks, PPE, and Facial Cues

Due to the spread of the COVID-19 pandemic in early 2020, the U.S. Centers for Disease Control and Prevention (CDC) implored individuals to regularly use disposable medical or reusable cloth masks when interacting in social public environments to mitigate pandemic contagion [35]. This sudden increase in mask usage also stimulated creativity and innovation and making and wearing masks for PPE [112]. Research into mask safety is ongoing with novel developments and innovations emerging at the time of this research [184]. Previous work on masks focused primarily on
health considerations, material affordances, and design considerations. Given that daily garments often become the subject of embedded wearable innovation such as smartwatches and e-textiles, masks present a promising candidate as computational augmentation.

One consideration within this domain of mask design affordances is the obfuscation of the face incurred through mask usage. Prior work in the literature demonstrates that facial cues and non-verbal communication are vital to understanding verbal communication in social interactions [71]. Non-verbal communication is especially important for populations with dementia, who are deaf or hard of hearing (DHH), or have cognitive impairments [160]. Proliferation of mask usage due to the COVID-19 pandemic has made interpersonal communication increasingly difficult, due in large part to the obstruction of facial cues by mask usage [163]. While some work has explored social and behavioral means of mitigating these obstacles to interpersonal communication, there has been little investigation of innovating mask design to better facilitate non-verbal communication [203]. Bowman et al. demonstrated through their Looking Glass prototype the potential for enriching interpersonal communication through computationally augmented wearables [24]. Computational augmentation accounting for non-verbal communication is exhibited elsewhere in computer mediated communication. Emoticons and pictographs, for example, have been demonstrated to services as non-verbal communication modalities in computer-mediated technologies, as well as other textual means of nonverbal communication [144] [226]. Given this prior literature, it is evident that further investigation into the potential for computationally augmented wearables as a means of enriching interpersonal and nonverbal communication is needed.
4.2 Related Work

4.2.3. Augmenting the Face with Computation

Work within the realm of computationally augmenting the face is limited, and reflects many of the use cases evident in the greater field of wearables. Many face-worn computational devices provide medical services to users [250]. The face has also been investigated as a source of novel interaction capabilities, such as using the nose [91], ear [142], or tongue [178]. More closely related to our avenue of inquiry are systems which enable mixed-reality capability for face-worn wearables. Chameleon, for example, uses a large display worn over a person’s face to provide embodied telepresence interaction for a remote user [170]. The potential of this paradigm has also been studied within the realm of accessibility. Augmenting and alternative communication (AAC) systems provide services for facilitating communication interaction for users who might have difficulty producing speech or language. Feuston and Jackson, for example, used a projection mapping system to display facial expressions directly on user’s faces whom had facial paralysis, thus enabling a display of the user’s emotional affect [68]. Developments in computational masks have emerged within the research literature as well [175]. Recently, commercial products such as the C-mask, and research prototypes such as MAScreen have begun integrating speakers and LED arrays into masks to provide facilitate communication. While MAScreen is able to provide a semblance of facial expression, they are abstracted and displayed using a grid of LEDs. In addition, unlike PokerFace, prior literature is unable to display captions and facial expression at the same time.

4.2.4. Literature Gap

To summarize, prior research highlights the potential of imbuing daily face-worn wearables with computation. We explore mixed-reality affordances to facilitate communication and mitigate the interaction barriers endemic to the mask form-factor. Building a working prototype enables investigating how it could potentially enhance
4.2 Related Work

PokerFace communication, interaction, and collaboration and end-user reactions and feedback to the affordances presented by this design space.

To explore the potential for computationally augmented masks to better facilitate communication, we designed and implemented a mixed reality prototype mask called PokerFace. Informed by prior literature, our design accounts for the loss of facial-cue and other non-verbal communication through textual and visual mixed-reality augmentation [71]. Captions and future augmentations, in particular, present a promising medium for facilitating the non-verbal interpersonal communication obfuscated by mask form-factor [144]. Our prototype PokerFace mask uses recyclable cardboard material and a commodity smartphone to embed computation into a helmet-like mask form-factor. This design allows a live video stream of the user’s mouth to be displayed on the mask’s surface, rendering the user’s face visible while keeping it covered and sealed by the mask’s facepiece. Additionally, captioning of the user’s speech can be displayed at the bottom of the video feed in real-time.

4.2.5. Implementation

PokerFace supports mixed reality interactions by embedding a smartphone in the mask’s facepiece. The infrastructure of the mask was created using recycled cardboard to architect sufficient support for the smartphone, while cradling the facepiece a sufficient distance from the face (Figure 4.2a). Double-ply cotton cloth was then used to seal edges of the mask around the nose and chin to prevent potential leakage of particles from the mouth cavity. To accommodate the offset position of the built-in camera on the smartphone, we also used a commodity endoscopic camera which attached to the USB-C port on the smartphone (Figure 4.2b). To display captions, either with or without video, we used the Microsoft Teams video calling platform with its real-time speech-to-text captioning service.
Section 4.3

USER EVALUATION

To understand how the PokerFace prototype facilitates communication compared with traditional masks, we recruited 18 participants for a user study. The study comprised an entrance questionnaire, a 4-condition counterbalanced evaluation of our prototype, and an exit questionnaire and interview. Our study design was reviewed and approved by our institution’s Ethics Review Board, and took place over the duration of 1 week.

4.3.1. Method

The 13 question online entrance questionnaire collected demographic information including age, occupation, self-described disabilities (if any), and gender, as well as surveying the user’s previous experience wearing masks for PPE. Participants were then given a virtual tutorial via video conferencing software on playing Guess What, a charades-like game for mobile devices. Guess What is played by one player, called the guesser, placing a mobile phone on their forehead while it displays a picture of a noun which has not been seen by the guesser. The other player, called the clue giver, gives
4.3 User Evaluation

PokerFace

clues without saying the actual word displayed to get the guesser to say the noun. Upon correctly guessing the noun, the guesser tilts the phone down to display a new word on the mobile device. Players can “pass” at any time by tilting up if the word proves to be too difficult. The number of nouns the clue giver is able to communicate to the guesser within 90 seconds is tabulated as the score. The tutorial detailed how to play the game, and a practice round where both participant and research team member were the clue giver. The study involved a series of 8 games of Guess What under 4 game conditions: traditional mask, PokerFace prototype mask with video, PokerFace prototype mask with video and captions, and PokerFace prototype mask with captions and without video. To protect the health of our participants under pandemic conditions, it was imperative that we limited the amount of mask removal and mask-changing which they performed during the study. Thus, we employed the use of a researcher as a confederate, who served as the playing partner for each participant, and was the only player to change their masks under each condition. For each of the 4 conditions, the participant’s mask did not change, wearing their personal cloth or disposable traditional mask for PPE throughout the duration of the study, while the confederate changed their masks for each of the 4 conditions. Once the confederate had changed into 1 of the 4 masks as dictated by the condition, they played 2 games of Guess What with the participant, one game where the confederate served as the guesser and the participant served as the clue giver and a second game with these roles reversed. We only allowed the participant to “pass” on any noun in a game, since the confederate became quite proficient with the game. Between conditions, the confederate changed their mask (a safe distance away) according to the next condition, and again played 2 games as before. After completing the 8 games of Guess What under 4 conditions between the participant and the confederate, participants completed an online exit questionnaire followed by an online exit interview. The exit
questionnaire comprised 12 short-answer and 20 Likert Scale questions asking participants to rate their experience with various aspects of interacting with and wearing traditional masks, as well as experiences interacting with the prototype mask during the study. When possible, questions involving specific aspects of interacting with people wearing masks were repeated using the same language for both traditional masks, as well as the PokerFace prototype mask. This was done to afford a Likert scale basis of directly comparing participant experiences with these mask features. The exit questionnaire also asked 16 short-answer questions (total of 36 questions) reflecting on their experiences with the 4 mask conditions. Semi-structured exit interviews directly followed completion of the exit questionnaire, giving our researcher an opportunity to review participant’s responses to the exit questionnaire and clarify any vagueness or inquire about inconsistencies.

4.3.2. Participants and Data Collection

All 18 participants ranged in age from 18 to 84 and had previous experience using masks for PPE. Study participation was strictly voluntary, taking place over a duration of approximately 90 minutes, and all participants received a $50 gratuity for their time. Eight participants (44%) identified as “woman”, eight (44%) as “man”, one (6%) as “genderqueer”, and one (6%) preferred not to say, and 2 participants identified that they are DHH. All of our participants indicated that they had previously played Guess What or similar Charades-like game. We collected quantitative data in the form of the number nouns guessed correctly per game (score). During the exit questionnaire, Likert scores on a scale of 1 to 5 were collected with 1 meaning “not at all” and 5 meaning “very much so”. Qualitative data were accumulated during both the entrance and exit questionnaire, as well as the exit interviews, and recordings of the game sessions. These participant responses were selectively reviewed by 2 members of the research team who devised an overall coding approach using a
grounded theory methodology. One of these two researchers then applied this agreed-upon coding approach to the remainder of the participant responses. Qualitative responses were coded for recurring themes in participant experience interacting with the confederate under the 4 study conditions. This analysis, as well as descriptive statistical treatment of the quantitative and Likert responses are detailed below.

![Comparing Scores of Confed and Participant as Guesser](image)

Figure 4.3: Results of user study comparing confederate and participants scores, reflecting the same trend across both datasets.

**Section 4.4 RESULTS**

We first analyze the results of our quantitative data reflecting participant overall performance interacting with various features of the PokerFace prototype versus traditional masks. Then we look at the qualitative data collected in the form of Likert results and open-ended questions from our entrance and exit questionnaires. After examining the limitations of our design approach, we proceed to a discussion of
these results contextualized by the feedback provided by our participants in the exit interviews and questionnaires.

4.4.1. Participant Task Performance

Overall, our participants performed better at correctly guessing clues with any version of the prototype PokerFace mask compared to traditional masks (Table 4.1). The performance score was similar under the video without captions ($\mu = 12.2$, $\sigma = 3.0$) and video with captions condition ($\mu = 11.6$, $\sigma = 3.2$). Interestingly, participants performed better using the prototype mask with captions but without video ($\mu = 9.5$, $\sigma = 2.4$), than with traditional masks ($\mu = 7.9$, $\sigma = 2.0$). These findings suggest that the video of the mouth contributed more toward increased performance than the captions. Due to the small dataset of users (n=18), standard deviation for the reported results was rather wide (Figure 4.3).

4.4.2. Qualitative Likert Feedback

We also collected Likert scale results of participant experiences as part of our exit questionnaire. Participants were asked to rate various aspects of their experience
interacting with the prototype, as well as traditional masks on a scale of 1 to 5 where 1 indicated “not at all” and 5 indicated “very much so”. Figure 4.4 shows a selected subset of Likert scale questions that compare reactions to traditional masks with the prototype. Overall, participants reported that communication was easier when wearing the PokerFace mask than traditional masks, with 90% of participants reporting a score of 4 or 5. When asked this same question in regard to traditional masks, 22% of participants reported a score of 4, with none reporting 5. Other general trends evident in the Likert results indicate that participants generally do not enjoy wearing traditional masks, with 65% reporting 2 or below when asked if they “liked wearing a traditional mask” and none reported a score of 4 or 5. This is echoed by the 78% of participants who reported that traditional masks were not fun to use (rating them a 1 or 2). In comparison, most participants also found the PokerFace mask “fun” to interact with, reporting a score of 4 or 5 for 75% of participants, and none indicating a score of 1 or 2.

4.4.3. Limitations of Our Results

Our 4-condition study design did not achieve the 24 participants needed for a full Latin Square counterbalanced study accounting for all possible permutations of condition ordering. Our principal concern in orchestrating our study was on the safety and health of our participants and researcher. Using a confederate as partner for all games played during the study introduces limitations to the data we collected as well as potential biases in our reported results. Since we used the same confederate across all conditions and orders, there is the potential that our confederate could improve task performance over time due to learning effects. To account for this potential bias, we conducted an analysis averaging confederate and participant performance over time per trial (Figure 4.5).

Figure 4.5 shows that both confederate and participant guesses remained compa-
Figure 4.4: Results of Likert Scale questions from exit questionnaire and interviews.
Figure 4.5: Averaged results by trial reflecting comparable performance by participants and confederate over time.

rable across all trials, with a slight improvement for both confederate and participant guessing over time. Note that the data also reflect that participants consistently performed better at guessing (on average) than the confederate. A similar per-condition analysis above that compared performance of participant and confederate for each condition (Figure 4.3) also found that the participant consistently performed better at guessing than the confederate across all conditions. These data confirm that the task remained meaningfully challenging over time, perhaps in part due to the mitigation of not allowing the confederate to pass on difficult clues. This indicates to us that influences of learning and order effects are fairly minimal. While acknowledging the limitations of our quantitative data, we emphasize that we relied on our qualitative and Likert results to understand the interaction experience with PokerFace. Furthermore, our participants were only able to reflect upon experiences interacting with the confederate wearing the prototype, and not wearing it themselves. However, given that the main benefits of the mask are experienced by those interacting with the mask wearer, and not in wearing the mask itself, reported participant experiences interacting with the mask are valuable to understand. Analysis of this is crucial to
inform the design of future face-worn wearables and understand communication while wearing these devices.

Section 4.5

**DISCUSSION**

While general trends indicate that participants preferred the prototype mask because of its ability to better facilitate communication, these results were not unanimous. Participants indicated several social stressors which should be considered for future XR mask designs. Users also provided feedback specifying use-cases where they thought the prototype would be useful in their work or social interactions. We discuss our prototype’s perceived effect on communication, social awkwardness and wearability, and participant-reported use case scenarios.

4.5.1. XR Masks and Communication

Participants overall perceived communication to be easier when interacting with the PokerFace mask than when interacting with a person wearing a traditional mask, as also corroborated by the quantitative performance results. This perceptual and quantifiable improvement to communication can be attributed to two principal affordances enabled by the prototype: facial visibility and speech captioning.

*Face Visibility.* Quantitative analysis of participant performance indicated that using the PokerFace prototype mask with video tended toward better performance at the collaborative communication game. Similarly, most participants indicated that the principal benefit to communication afforded by the prototype is the ability to see their communication partner’s face. Participants reported that being able to read facial cues from the PokerFace video display was extremely helpful in understanding the wearer’s intent. “The interaction was easier and more engaging when wearing
the prototype. Normally with a traditional mask I have to get closer than the 6ft minimum to hear what the other person is saying if they are a senior citizen or someone who is just soft spoken. I didn’t realize how helpful facial cues are when talking. The prototype mask was a fun way to engage with the other person and see their mouth … without removing the mask and feeling at risk of a virus.” (p9) Face visibility was almost unanimously indicated by our participants as being incredibly helpful in communicating, with many echoing p9’s sentiment about reading facial cues. Including p9, fourteen participants (78%) volunteered that they noticed themselves reading their game partner’s lips when visible through the live video feed. While two (p17 and p18) of these participants disclosed they were DHH, the others did not, suggesting that lip and facial cue reading is a useful affordance of the mask for all users. Even when not addressing the mask directly, participants indicated desires to see other’s faces. “I’ll be glad when things go back to normal, if they ever do. I miss seeing peoples smiles without the masks.” (p12) The ability to perceive emotional cues and facial expressions was a priority among numerous participants, especially those with highly social occupations. Participants who worked as educators (p8, p9, p11, p15) expressed a need to have their and their students’ mouths be visible when teaching. Other participants who held social occupations indicated similar priorities with being able to see other people’s faces when performing their work. For example, p1 is a beauty consultant, reflected that being able to see their customer’s face was imperative for making successful suggestions, as well as their customers being able to see their smiling face to establish rapport. Furthermore, 89% of participants indicated in the exit questionnaire that the live-video feed was helpful in facilitating communication, and thus was a promising feature of our mixed-reality prototype.

**Captions.** The speech-to-text captions displayed in real-time on the PokerFace screen was the second integral element in facilitating communication and improv-
ing performance in our study task. The better performance of the PokerFace mask with captions but without video condition seems to indicate an intrinsic benefit of captions independent of a live video display. All saw the potential for captions to be useful, and expressed preferences for how the captions should be displayed. Interestingly, some participants (p4, p6, p11, p12) indicated they preferred the captions without the video feed because the captions with the video were overwhelming.

"I'd actually prefer [captions without video]. Only because reading body language, listening to my partners voice, reading captions as well as looking at the screen of the other phone was more than enough to play charades. Introducing video was information overload." (p12) Other participants indicated that they found captions with video more useful because they could use the captions to confirm what was being said. Two participants (p13, p15) compared this experience to “watching movies with subtitles”. Our participants who are DHH reported the most enthusiastic support for captions in the qualitative feedback, indicating that the captions often filled-in words that the person may have missed when someone is speaking. “I sometimes don’t hear what said because the mask covers people’s mouths. So, I just nod my head like I hear them...I don’t want to be rude and say I missed something. So [seeing the captions displayed] was nice because I could check the screen if I didn’t hear [the speaker].” (p18) Both of our participants who are DHH preferred video with captions over video without captions. One DHH participant reported a reduction in stress when they stopped trying to hear the speaker, and instead focused on reading the captions and lips of the speaker wearing PokerFace. These results indicate that captions are a promising accessibility feature for a mixed reality face-worn display.

4.5.2. Mask Form Factor and Social Interaction

Participants reflected a notable period of adjustment when interacting with a PokerFace mask wearer, citing a “social awkwardness” (p15) or an “uncanny valley effect”
4.5 Discussion

(p11) that subsided after continued interaction with the device. As mentioned above in Section 5.2, participants indicated an incongruity between their feelings of silliness when wearing a mask in public, and their feelings of silliness when others wear masks in public. When participants were asked if they would use PokerFace if available, 50% said they would. Participants who were skeptical about adopting PokerFace often held highly active, or less social positions, and their feedback reflected that the device would be less useful to them in these circumstances. When asked if they would feel silly wearing the PokerFace mask in public 50% of participants indicated that they would feel silly, while 73% of participants indicated that they would not feel silly if someone around them was wearing a PokerFace prototype mask. This reaction was contextually dependent, and even participants who anticipated feeling silly while wearing PokerFace in public, still delineated several scenarios where using the prototype in public could be beneficial. For example, several participants felt the novelty of the prototype was “fun” and would be jovial to wear to a party (p3, p4, p6, p8, p11, p12). Others identified use case scenarios where communicating may be difficult, such as trying to speak with friends at a location with substantial background noise, and thus the prototype could be beneficial for enabling communication (p1, p16, p17). These user-reported use cases, in combination with the feedback from participants who were enthusiastic about the potential of PokerFace, seem to indicate that all participants acknowledge that the device would be beneficial in facilitating communication in social circumstances.

The mask form factor became a topic of hesitation for participants as well, several of whom noted that the current prototype was “too bulky” (p8) for regular usage or “wouldn’t be able to fit in my purse” (p1). While future iterations of PokerFace could substantially reduce the overall bulkiness of its form factor, this concern from participants reflects a design priority for portability among participants. Subsequent
designs of mixed reality face-worn wearables could make use of flexible displays to substantially reduce the overall size and weight of the device. Weighing the trade-off cost of portability and durability of material is a design consideration that should be incorporated into the future creation of mixed-reality masks.

Section 4.6
MIXED-REALITY MASK APPLICATIONS

Motivated by our study results and participant feedback, we describe potential applications and interactions enabled by mixed-reality masks such as PokerFace. Many of these interactions were provided by our participants during the user study, or were indicated as potentially useful applications extracted from qualitative user feedback analysis. While some of these use cases could be enabled by a simpler, transparent mask that enabled viewing the mask-wearer’s mouth, most rely on features that could only be enabled by digitally augmenting the mask. We acknowledge that the size and form factor of the current PokerFace prototype make it impractically complex, we believe that our study demonstrates the opportunities in this design space. In light of the lightweight, bendable displays and advances in generating avatar re-enactments of bodily expressions obscured by wearable devices [42], we believe that a practical prototype that affords the features of PokerFace could be implemented in the near future.

A common application suggested by our participants was communication with people who are DHH. This suggestion was particularly emphatic from our two participants who are DHH. Affording lip reading and presenting live captioning could enable more fluid communication for users who may be DHH. The built-in microphone and speaker on most commodity smartphones could be used to amplify the voice of the person speaking while wearing the PokerFace prototype, overcoming the muffling
effect of traditional masks. Participants who were not DHH also identified additional
communication use cases with these features, indicating that widespread adoption
of an accessibility device with these considerations could be possible. Several partic-
ipants worked in environments which were heavily bilingual, but were not bilingual
themselves, often relying on translation services. Overwhelmingly, integration of live
translation into future iterations of the PokerFace prototype or other mixed-reality
masks was indicated as a valuable feature by these participants working in bilingual
professions. Many real-time translation services currently exist, and integration of
this feature into PokerFace, combined with the live video feed, could enable a user
to communicate in a foreign language while the expression of their mouth remained
visible. XR Filtering such as those used in Snapchat could enable a variety of aug-
mented communication capabilities, from personal expression and creative pursuits,
to serving as an integrated element of a costume. This use case seemed particularly
popular with users who often offered whimsical applications of the prototype in their
feedback. For example, p10 asked “Is there any way to make a mask with the face
of Mickey Mouse smoking a fake cigarette attached to it?” Similarly, users with fa-
cial paralysis [68] could potentially benefit from having augmentation filters which
facilitate exhibiting desired facial expressions or restoring facial symmetry.

Section 4.7

CONCLUSION

The PokerFace prototype explored how a mixed reality mask could overcome com-
munication limitations of traditional face masks that became widely used during the
pandemic. Our user study explored the various features of our prototype PokerFace
mask, and compared the effectiveness of these features at allowing players to commu-
nicate while playing a collaborative communication game. On average, participants
performed better using any of the 3 PokerFace conditions than with traditional masks. Feedback from participants indicated preferences for various features of PokerFace, as well as suggested potential use cases. Some participants preferred to use PokerFace with captions and without live video because it mitigated communication difficulties arising from voice muffling, without inducing information overload. Participants who worked in social disciplines such as education were more enthusiastic about the potential of PokerFace, while all participants, even those skeptical of our mixed reality mask, suggested social use cases where PokerFace could be beneficial. Social situations such as communicating with a person who is DHH, in environments which are loud, or in a foreign language were all indicated as potential communication use cases which could benefit from mixed reality masks. PokerFace is a preliminary work that highlights a promising new design space that could provide practical solutions to enable and facilitate better collaboration, as well as prove fruitful ground for novel XR interactions. We believe the potential posed by integrating mixed reality into masks merits future research even beyond the use of masks during a pandemic.
Co-creative AI tools provide a method of creative collaboration between a user and machine. One form of co-creative AI called generative design requires the user to input design parameters and wait substantial periods of time while the system...
computes design solutions. We explore this interaction dynamic by providing an embodied experience in VR. Calliope is a virtual reality (VR) system that enables users to explore and manipulate generative design solutions in real time. Calliope accounts for the typical idle times in the generative design process by using a virtual environment to encourage parallelized and embodied data-exploration and synthesis, while maintaining a tight human-in-the-loop collaboration with the underlying algorithms. In this paper we discuss design considerations informed by formative studies with generative designers and artists and provide design guidelines to aid others in the development of co-creative AI systems in virtual environments.

Figure 5.2: Conceptual Illustration of the Calliope system featuring a collection of objects generated during design walkthrough

---

**Section 5.1**

**INTRODUCTION**

Generative design [3] [14] [15] [247] [235] presents a promising opportunity for co-creative AI that leverages the computational power of design simulation, parametric design, optimization techniques and artificial intelligence, enabling computational
tools to play an active, participatory role in the design process. Generally, this technique enables designers to explore a greater number of design solutions than traditional 3D modeling processes, or generate solutions that would be difficult or non-obvious for humans to create alone [57] [187]. However, the typical workflow of these interfaces is antithetical to the mental model of traditional design tools. Often, generative design software obligates the user to parameterize the problem and solution space up-front and then requires the user to wait a substantial period of time while generating candidate solutions. Aside from being inefficient, this workflow is counter-intuitive to the mental-model of the creative process, which is iterative, extemporaneous, and playful, often employing a divergent problem-solving process, as opposed to the convergent process of generative design. Furthermore, by encouraging the user to specify all constraints of their problem up-front, exploratory and serendipitous browsing are eliminated as potential interaction methods, thus resulting in the user primitively eliminating many potential and inspiring solutions from their design space before fully exploring the problem [83]. While prior work has explored embodying the design process to locate the desired design solution (e.g. get the right design) [254] [9] [253] our work embodies the design exploration process in order to create the right design [232].

Prior work in this domain has largely focused on 2D interfaces for procuring design solutions. DreamSketch, for example, supported greater design creativity in generative design workflows by enabling a user to input design constraints via sketching [110]. However, this workflow may not be optimal when working in 3D, since traditional generative design systems for 3D objects rely on 2D graphic interfaces [187] [111]. Virtual Reality (VR) offers a promising platform that allows designers to sculpt and model 3D objects using methods more attune to traditional sculpting tools, as well as supporting embodied exploration of data. Furthermore, VR offers a more engaging
experience, which can enrich the synthesis experience as well as mitigate inactivity due to algorithmic idle times. However, this modality has largely gone unexplored for generative design. In this paper, we present Calliope (named for the Grecian muse of architecture and epic poetry) an interactive model synthesis and exploration tool that helps designers explore divergent solutions to a given design problem within a virtual environment. The system allows users to inspect the visual appearance of geometric models, directly edit their geometry, and receive design input from a generative adversarial neural network (GAN). By leveraging the spatial exploratory potential of virtual reality, our system keeps the user engaged in synthesizing and evaluating other geometric models while the GAN computes potential solutions to the designer’s problem. This approach facilitates an iterative design process, while encouraging a collaborative dialogue between the machine and designer. Similar ideas have been explored in the realm of literature by Borges, however, Calliope comprises the first human-AI creative collaboration interface for VR. The contributions of this work include:

- an exploration of the design considerations implicit in creating VR interfaces for generative design;

- A series of interaction techniques to allowing designers to create, combine, and explore divergent solutions to a given design problem within a virtual environment, working in close parallel with the underlying algorithms;

- a proof-of-concept system, Calliope, which enables designers to work closely with underlying generative algorithms to creative exploration within an engaging VR environment;

- insights for designing human-AI creative collaboration interfaces in VR extracted from multiple design validation sessions.
This work builds upon creative authoring platforms in virtual reality, co-creative and mixed-initiative design, as well as rapid-iteration and collaborative AI interfaces.

### 5.2.1. Creativity Support in VR

Prior work in VR literature investigates the potential for immersive environments supporting 3D design tasks such as sketching and modeling [202] [254] [8]. More recent inquiries explore the mechanics of sketching and modeling in mid-air, emphasizing the integrity and quality of the input gestures [9] or suggesting improvements and corrections to the design [12], as well as enabling more advanced workflows. Similarly, DreamRooms, allows a user to interact with a procedural system to develop 3D room layouts for VR [253]. Many initiatives exist in mixed/augmented (MR/AR) reality as well. MixFab, for example, allows users to 3D model objects for fabrication using an AR gestural approach [252]. Most closely related to our interests is Mix&Match, which enables users to sample and view artifacts from Thingiverse in AR within the context of their physical environment [222]. Our work distinguishes itself from these in that it allows user to work directly with generative design algorithms in a collaborative process and enables creative exploration earlier in the design process to find the right design.

### 5.2.2. Co-Creative AI and Generative Design

User-driven generative design tools enable a user to specify high-level design intents, and systematically produce candidate design solutions using generative algorithms. Previous systems have applied this approach to optimizing office building layouts, fabricating airplane partitions, and designing furniture [236]. The pipeline in these
systems automate parts of the design process by enabling users to specify abstract design goals and high-level constraints to the system which are used to produce candidate design solutions [37]. Tightly-looped interactivity between the user and the underlying algorithms in these systems is often difficult for a variety of reasons, such as the high latency of the algorithms themselves. Typically, after specifying high-level design goals, the designer must wait hours or days before viewing the results of the algorithmic generation process. Previous efforts have attempted to address this issue such as eifForm, a generative design tool combining structural model generations with a traditional modeling approach [211]. Martinez et al. enables designers to input sample patterns and obtain a design optimized for structural integrity and aesthetic similarity to the input [151]. While promising, these prior works do not permit users to directly manipulate system output and provide feedback on desired solutions, nor keep the user engaged while the system produces candidate designs. Furthermore, even after waiting substantial time to view results, the designer then has to manually sort through all the design candidates generated by the algorithm. While these algorithms are effective at producing a multitude of solutions, sorting through these solutions for ideal designs still requires substantial work on behalf of the designer. To account for these limitations, recent work has investigated generative design interfaces which explore and visualize large sets of data and design solutions [152], as well as systems which interprets sketches created by the user as problem representations [37] [110]. Most closely related to our work is Forte, which allows a user to iteratively design 3D printable objects using interaction techniques designed for topology optimization [37]. Early explorations of embodying the collaboration between human and machine seem promising [145][162][137] but have yet to extend this domain of inquiry to the embodied system of VR.

Our work distinguishes itself from these predecessors by leveraging the spatial
properties of VR to enable users to interactively explore design artifacts and specify desired design features to the underlying algorithms. Calliope enables designers to interact with the generative process in a tightly looped human-ai interaction by encouraging users to explore candidate designs in parallel. Additionally, Calliope employs the interactive modalities of VR to encourage embodied exploration and synthesis.

5.2.3. Collaborative AI Interfaces

Mixed-initiative interfaces support collaboration between an intelligent system and a designer, allowing both the user and the intelligent agent to “do what they do best” by assuming different, but complementary roles in the design process [34]. Within the scope of design tools, user-driven suggestive interfaces are mixed-initiative interfaces which enables users to control and drive the design task. While the user controls this process, the intelligent system observes, analyzes and suggests improvements or alternatives to improve the design. An early work in this field is Chateau, a 3D sketching tool which predicts what a designer will draw next and suggests alternative completions to the drawing based on observations of user input [100]. Other examples include the work of Umetani et al. who presented a furniture design tool which indicates unstable structures while a user edits their design, and offers alternative suggestions for solving these instability problems [236]. Additionally, Tsang et al. developed a system which takes in target design images as input, generates 3D curves, observes users’ input strokes, and suggests relevant geometry accordingly [234]. In contrast to this previous research, Calliope enables designers to perform direct mesh manipulation on generated 3D artifacts in a virtual environment and provides design candidates based on features specified this way. Further, Calliope leverages a generative approach to explore a multitude of design candidates within a VR environment.
Section 5.3

FORMATIVE INTERVIEWS

To better understand the role that embodied experience could play in dealing with delays, maintaining engagement, and exploring design variations we conducted a series of semi-structured interviews with end-users of generative design applications.

5.3.1. Participants and Methodology

We recruited 6 participants (3 female identifying, 3 male identifying) between ages of 22 and 43 for a remote semi-structured interview. Participants were familiar with generative design through their creative practice or professional development experience. 5 participants were familiar with VR/AR in some capacity and 2 of these had VR development experience. The interview consisted of demographic questions and 11 open-ended interview questions discussing prior generative design, human-AI collaboration, and experiences designing in VR/AR as well as their overall experience with these mediums.
5.3.2. Design Considerations and Goals

Guided by prior literature and the results of our formative interviews, we performed an iterative qualitative analysis of the transcribed participant interviews informed by prior ground-theory research best-practices [222]. Following the interviews, an author coded the transcribed data for need-finding design considerations and participant expectations for interaction with such a system. From this, we extracted a series of design goals and considerations for our system outlined below.

**Tightly-Looped Human-AI Collaboration (D1).** The traditional workflow of generative design is a linear, solution driven process that requires the designer to input design requirements up-front, and then wait substantial periods of time to generate solutions. These solutions then must be manually explored which is an additional time-consuming and tedious process. P6: “It can be tedious... you want to do a thorough job and not miss anything cool but you also get tired after looking at [so many] things”. Any generative design system should expedite this process, allowing a user to efficiently explore many design decisions, and support a creative dialogue between the user and the algorithm. This allows a user to freely explore and test as many design options as possible, assisting in getting “the design right and the right design” [232] [232].

**System Transparency (D2).** All of our participants discussed an underlying uncertainty with underlying generative algorithms. P3: “It can be exciting to not know what the machine will dream-up... but it can also be frustrating... I sometimes feel the machine knows more about [my design] than me.” Furthermore, many of our participants mentioned potential distrust when using such systems P1: “I’ve crashed [the generative system] so many times I’ve lost count... I never understood what decisions will break the software.” Evidence suggests that explainability of the underlying AI
system is crucial for fostering trust with users [62]. We believe that an interface for these systems should mitigate these concerns by providing as much transparency as possible.

**Creativity vs Creation (D3).** When discussing current limitations to creativity in VR, we found a distinction between creation (the physical acts of creating an object) and creativity (scaling the act of creation to allow exploration of many ideas). Many participants indicated that they believed quality of ideas came from quantity, and thus creativity encouraged the generation of as many design ideas as possible, sorted and explored to find quality candidates. P2: “My old boss used to call it the ‘blah blah blah gold’ philosophy. You have to create a lot of blah to get gold”. However, current VR sculpting practices are limited by the difficulty of creation in VR. P5: “It just takes forever, and by the time I’m halfway done [sculpting] something real basic I start to feel a little sick”. Thus, we believe it is imperative that generative systems support rapid creation and ideation to support creativity within VR.

**Leverage Spatial Interaction for Engagement (D4).** One common obstacle to tightly-looped human-AI interaction involves the high-latency computation cost, and thus the resulting speed inhibition, of generative design [37]. P1: “I’ll send samples off for processing which can sometimes take hours and the results that come back are garbage.” We believe a system that utilizes VR must leverage the engaging properties of the virtual modality to accommodate for this limitation. Furthermore, many generative design systems are intended to create 3D objects, yet rely on 2D means of interaction, which is counter-intuitive for users. P6: “Sometimes while modeling I get frustrated looking for the right tool... all I want to do is reach into the screen and pinch sculpt the mesh”. The spatial nature of VR not only provides affordances to keep a user engaged during computation idle times, but also supports
a more intuitive process for 3D modeling [231]. Thus, any such system should aim to keep the user engaged exploring and synthesizing additional design solutions while the system computes solutions in parallel.

Section 5.4
DESIGN OF IMMERSIVE CO-CREATIVE AI SYSTEM CALLIOPE

To account for the above design considerations, we developed Calliope: a user-driven interface for tight-looped human-AI creative collaboration within a virtual environment [Figure 5.3]. While system development was largely guided by our formative interviews and literature review, we acknowledge that this is only one possible instantiation of the higher-level concepts and design goals. This system can be considered an initial exploration to guide future development of co-creative AI in VR.

5.4.1. Virtual Environment

The structure of our virtual environment drew inspiration from radial data visualization and is progressively populated as it is explored by the user. Given that the designs produced by Calliope lend themselves to organization that is both hierarchical and categorical, the results are well suited for representation in a pseudo radial data visualization [89]. Furthermore, this representation translates easily to VR, and mimics the structure of the labyrinth from Borges’ Las Ruinas Circulares, which helped inspired this work. Users are required to develop a certain number of objects within each room before additional rooms are spawned. This was designed to encourage the user to explore additional design options before continuing (D3). The placement of the room is dependent upon the kind of modification and generation performed by the user in their current room. This methodology also allows the user to construct a
5.4 Calliope Co-Creative AI

Figure 5.4: The Calliope Virtual Environment Design Was Informed By Radial Data Visualization.

visual representation of their generated artifact’s lineage as they proceed through the system. The rooms themselves are color-coded in order to provide the user a sense of variation in room design, assist with a user’s memory of the environment, as well as further support the visualization of their design process [89]. Each room, once spawned, is populated with a number of vitrines determined by the actions in the previous room, which serve as the central locus for providing input and receiving output from Calliope [Figure 5.6]. Vitrines were chosen as an interactive locus in order to evoke a metaphor between the synthesis of data and the creation of a museum or gallery exhibition. This way, the user develops an exhibition of iteratively designed artifacts while procedurally constructing a visualization of their design interests.
Here, we provide a walkthrough demonstrating the design process using our human-AI collaborative system. In this example, the designer would like to create a custom chair for a client that meets specifications such as height, material, and presence of certain features such as armrests and head rests.

5.5.1. Initializing Objects

The structure of the virtual environment drew inspiration from radial data visualization and is progressively populated as it is explored by the user [Figure 5.4]. The designer begins by selecting an empty vitrine [Figure 5.6] to begin the generation process. A menu appears with several options for generating an object [Figure 2B]. The user can select “sculpt” to sculpt a new mesh from scratch using a brush-like interaction or select “generate”. By selecting generate, the generation menu appears with a variety of object classes from which the user can select. The user selects “chair”, and a chair appears inside the previously selected empty vitrine [Figure 5.3C]. The user then proceeds to repeat this process with each of the vitrines in the room. Once a user has filled all the empty vitrines in the current room, a new room is spawned, containing new vitrines corresponding to the operation performed in the previous room. Each new room will contain a minimum of 2 empty vitrines to allow generation of new artifacts. If the user performed direct mesh-manipulation in the previous room, 4 additional vitrines will be spawned in the new room containing the results of mesh-manipulation interpretation. Similarly, if the user performed mutation in the previous room, 2 additional vitrines will be spawned in the new room containing the results of
the mutation.

5.5.2. Direct Mesh Manipulation
The designer selects an artifact containing visual features which they find interesting, but wishes they had other specific characteristics such as armrests (D1). To communicate this to Calliope, the designer selects the target object, and then chooses “edit” from the menu [Figure 5.3E]. The designer then uses the direct mesh editing raycasting tools to quickly sculpt an impression of armrests on their target object. These techniques are performed using a traditional ray-casting point-and-click interaction common to other sculpting platforms such as TiltBrush. Once the user has completed their mesh manipulation, a progress light appears, indicating that Calliope is performing a computational operation. Calliope will take the user’s manipulated object as input, and attempt to create similar designs to the user’s custom input. During this time, a user may venture to other rooms and conduct parallel design processes on other meshes or generate new objects. When Calliope has completed the computation, results are displayed to the user in the adjoining room.

5.5.3. Mutation
The user favors certain characteristics of one artifact, and certain characteristics of a different artifact (D2). Wishing to combine the two artifacts in order to render a single object containing a mix of the favored features, the designer uses the mutate function of Calliope. The designer selects a generated object inside a vitrine and selects “Mutate” from the menu [Figure 5.5]. When “Mutate” is selected, the user is prompted to select how many sampled objects they wish Calliope to generate (D3). The upper bound of generated objects was determined through our stress test in Section 7.3 to not overwhelm the system and induce frame skipping. The user indicates they desire 4 objects, and the walls of the environment are then lowered to
display all objects generated during the session, as well as the 4 randomly sampled inspiration objects generated by Calliope. The designer then selects the second object containing favorable features, and the environment walls are raised. A progress light displays inside the selected vitrine. After completing the requested interpolation, Calliope displays 4 mutated objects in an adjoining room [Figure 5.3H].

5.5.4. Parallel Design ideation

Rather than wait for Calliope to complete the computation, the user instead turns their attention to other empty vitrines inside the virtual environment (D4). They select the nearest door, and proceed to explore other rooms, selecting empty vitrines to generate new artifacts for consideration. Some of these artifacts contain desirable physical features which the designer favorites from the menu, saving these objects for future mutations. A light appears on the mini-map, indicating that Calliope has completed its computation. The user selects the newly generated room from the mini-map, teleporting the user to the room.

5.5.5. View Results

The user examines each result generated by Calliope during the mutation process. They decide to discard one result and favorite the other. The designer can then continue the design iteration process by generating and manipulating additional objects. After several iterations of generation, manipulation and interpolation, the designer returns to the central room and lowers the walls of the environment in order to survey all the objects created during the current session. During this time, the designer favorites their final design candidates which are saved to the working system as obj files for future use, fabrication, and reference. The design can be seen in Figure 5.9A.
5.5 System Walkthrough

Figure 5.5: Micro vs Macro Creation Viewing; A. User views a single design; B. User lowers walls to view all designs from this session.

5.5.6. Additional Interaction Techniques

In order to support creative collaboration between the designer and the generative algorithm within the virtual environment, we developed the following additional features and interaction techniques.

**Mini-Map and Teleportation.** In order to assist the user in navigating the system, a mini-map presents the user with a top-down view of the virtual environment, updated as the user progresses through the procedural generated environment [Figure 5.6A]. In addition, the mini-map informs the user of new rooms to explore, status of current Calliope computations, and location of favorite designs (D4). The user can also teleport to a specific room within the environment by clicking on a room within the mini-map.

**Micro vs Macro Creation Viewing.** In order to give the user a macro-view of all objects created, the walls of the environment can be lowered at any time, allowing the designer to view all objects they have created during the session (D3). The walls can then be raised in order to give the user a micro-view of the objects in their current room, allowing the designer to focus on current objects they are designing [Figure
5.5 System Walkthrough

Color-Coded System Indication. While Calliope is performing a given computation (either generation or mutation) we provide feedback on general progress as well as system status to the user in the form of a color-coded light system. Once the user has specified a given function for Calliope to perform, the vitrine is filled with a particle effect light, which glows brighter and moves quicker to indicate the progress of the given computation. These lights are also color coded to reflect the current GPU-memory consumption, since this is an adequate indicator of computational cost demanded of the system. If the GPU has currently consumed less than 40% of volatile memory, the lights remain blue, 40%-60% turns these lights yellow, and above 60% turns these lights red. These values were chosen because 60% and above memory consumption may affect system performance. This color-coded system subtly informs the user of system status and allows them to make informed decisions.
Figure 5.7: Color-Coded System Indicator communicates transparent system load status.

with subsequent interactions (D2).

**Mini-Map and Teleportation.** In order to assist the user in navigating the system, a mini-map presents the user with a top-down view of the virtual environment, updated as the user progresses through the procedural generated environment [Figure 5.6A]. In addition, the mini-map informs the user of new rooms to explore, status of current Calliope computations, and location of favorite designs (D4). The user can also teleport to a specific room within the environment by clicking on a room within the mini-map.

Section 5.6

EXAMPLE OBJECTS

We present the following sample design task walkthroughs in order to demonstrate the creative potential of Calliope. Our walkthrough tasks were completed by members of the research team and chosen in order to demonstrate the variability of design approaches supported by Calliope, as well as the application of our design considerations
Figure 5.8: Results of Design Session Indicating Breadth and Depth of User Creations mentioned previously.

5.6.1. Car Chassis
Similar to the described workflow, Calliope was used in the design of a car body which exhibited a series of unexpected visual characteristics [Figure 5.9B]. In this design session, the designer used a “breadth-first” design approach, asking Calliope to generate many sample designs. This process occupied half of the virtual environment, with one section dedicated to generating car designs, and another dedicated
5.6 Example Objects Co-Creative AI

Figure 5.9: Example Objects Generated Using Calliope: A. Custom chair; B. Car Body; C. Sofa Desk; D. Artwork

to generating other vehicle designs as inspiration. The designer then selects their favorite designs containing interesting visual characteristics and allocated an entire section of the virtual environment to interpolating these favorited objects with each other. Throughout this process, the designer didn’t feel the need to edit any of the objects directly, instead relying on Calliope to sample candidate designs and then mutate designs containing favorable visual properties.

5.6.2. Novel Furniture Form-Factor: Sofa-Desk
In this workflow, a user experiments with inter-class interpolation to explore candidate designs of novel furniture form-factors. The designer generates and discards many sample design candidates for sofas in a single vitrine until they find one exhibiting desirable qualities. In the empty vitrine next to it, the designer similarly generates and discards many sample candidates of desks.

After identifying two ideal candidates from two different classes, the designer mutates these two objects into the adjoining room. While the designer waits for this mutation to occur, they turn their attention to generating sample candidates in a different room, repeating the above process with different object classes. The designer repeats this process with many classes until they have mutated several quality candidates exhibiting desirable characteristics and novel form factors. The designer then performs mesh manipulation on these final candidates, in order to accentuate various qualities of the inter-class mutated candidates. This allows the designer to
fix any immediate structural flaws with the design candidate, such as missing legs or incomplete surfaces. Afterwards, the designer favorites their desired candidates for export as obj files [Figure 5.9C].

5.6.3. Sculpture
This session was significantly more unconstrained because the user embraced a highly experimental workflow. During this session, the user followed no discernible structure to their design process, instead choosing to follow their intuition completely. The user generates objects of various design classes and mutates them with each-other, performing mesh manipulation out of aesthetic experiment to “see what happens” more than functional directing the generative algorithm. Often, the user’s decisions are motivated more by a curiosity of the algorithms reaction to their decision, than any functional quality of the objects generated as a result of these decisions. The session concluded when the user arrived at an experimental result they deemed interesting, favoriting the result for export for possible fabrication [Figure 5.9D].

5.6.4. Video Game Scenery Suite
Generative systems often provide a narrow selection of solutions and produce a single design candidate. This is not ideal should a user require a collection of similar aesthetic objects. In this session, a designer wants to create an entire suite of scenery for a 3D video game. They allocate separate sections of the virtual environment for each type of scenery object they are modeling. After generating several samples of each individual class of objects, the designer selects favorites objects containing their desired visual aesthetics. The designer then mutates each favored object of a given class with the objects favored from other classes. This ensures that the final design candidates will contain visual characteristics similar to each other across different object class types. The designer then repeats this process with each remaining
class of object until they have created a collection of objects bearing similar aesthetic characteristics.

### Section 5.7 IMPLEMENTATION

In order to ensure replication of our system, we provide the following details on implementation and technical contribution. Our virtual environment was developed in Unity 2019.3.0f6 using OpenGL 4.5 on Ubuntu 18.04, and tested on Vive Pro VR headset. Our generative algorithm was tested on Pytorch 1.5 with CUDA 10.2 using python 3.8-dev on an NVIDIA 1080Ti GPU. Details on our generative algorithm are outlined below.

#### 5.7.1. 3D Generative Adversarial Network Terminology

Generative Adversarial Networks are appealing designers due to their promising ability to persistently generate novel objects [258]. Goodfellow et al. proposed the Generative Adversarial Network (GAN) which comprised two networks: a generator and a discriminator. The generator network synthesizes convincing objects in order to fool the discriminator [80]. Meanwhile, the discriminator attempts to distinguish between ground-truth objects (3D models taken from ShapeNet [36]) and objects synthesized by the generator. Training consists of the generator learning how to create 3D objects by adjusting weights corresponding to object features. Once trained, the resulting generator is able to produce, and interpolate between, the selected domain of objects taken from ShapeNet. We follow the architecture described by Wu et al. [258] to produce a generator which creates a representation of a 3D object by randomly sampling a z vector from a probabilistic latent space. This 200-dimensional latent z vector maps to a $64 \times 64 \times 64$ voxel cube, representing an object in 3D voxel space.
5.7 Implementation

The probabilistic latent space, in this case, refers to the solution space of possible objects generated by the system. Therefore, each \( z \) vector sampled from the latent space represents a novel object resulting from an interpolation of the 200 dimensions of the latent space. The discriminator \( D \) outputs a confidence value \( D(x) \) of whether a 3D object input \( x \) is similar enough to a ground-truth example to be considered acceptable or synthetic. The calculation of this classification loss is performed with binary cross entropy, thus the adversarial loss function can be described as

\[
L_{3D-GAN} = \log D(x) + \log(1 - D(G(z))),
\]

where \( x \) is a ground-truth sample object in a \( 64 \times 64 \times 64 \) space, and \( z \) is a randomly sampled noise vector from a distribution \( p(z) \). In this work, each dimension of \( z \) is an independent and identically distributed uniform distribution over each dimension of the latent space, in this case, represent a different geometrical aspect of the object. Generated objects are then rendered from voxel arrays to meshes using Marching Cubes and then refined using Laplacian smoothing before being imported into the virtual environment [147]. The above process will allow us to generate 3D virtual artifacts and render them as mesh objects by sampling a random latent vector \( z \) and mapping it to the space of 3D objects. We retain the \( z \) vector used for sampling our object for conditionally sampling the network after Mesh Manipulation or Mutation. To ensure we can conditionally sample from the GAN and reliably reproduce the same sample, we adopt the approach of Larsen et al. which proposed sharing the decoder of a Variational Auto-Encoder (VAE) with the generator of the GAN, resulting in a VAE-GAN [126].

The all-convolutional generator consists of five volumetric fully convolutional layers of kernel sizes \( 4 \times 4 \times 4 \) and strides 2, with batch normalization and ReLU layers added in between and a terminating Sigmoid layer. The discriminator uses Leaky
ReLU and otherwise mirrors to the generator. Because our workflow necessitates inferring latent vectors from observations (for example, if there exists a mapping from a label to the latent representation, we can then recover the 3D object corresponding to that label), we follow We et al. 3D-VAE-GAN [258] except embed labels instead of 2D images into the object space.

An additional encoder $E$ takes a label $x$ as input and outputs the latent representation vector $z$. The 3D-VAE-GAN network architecture consists of three components: an image encoder $E$, a decoder (the generator $G$ in 3D-GAN), and a discriminator $D$. The image encoder consists of five spatial convolution layers with kernel size 11,5,5,5,8 and strides 4,2,2,2,1, respectively. There are batch normalization and ReLU layers in between, and a sampler at the end to sample a 200 dimensional vector used by the 3D-GAN. We follow Larsen et al. loss function, which consists of three parts: an object reconstruction loss $L_{recon}$, a cross entropy loss $L_{3D-GAN}$ for 3D-GAN, and a Kullback-Leibler (KL) divergence loss restrict the distribution of the output of the encoder. Formally, these loss functions write as

$$L = L_{3D-GAN} + \alpha_1 L_{KL} + \alpha_2 L_{recon},$$

where $\alpha_1$ and $\alpha_2$ are reconstruction loss weights of the Kullback-Leibler divergence loss weight. From this we then have

$$L_{3D-GAN} = \log D(x) + \log(1 - D(G(z))),$$

$$L_{KL} = D_{KL}(q(z|y)||p(z)),$$

$$L_{recon} = ||G(E(y)) - x||_2,$$

where $x$ is a training sample 3D shape, $y$ is the training sample’s corresponding label, and $q(z|y)$ is the variational distribution of the latent representation $z$. The
KL-divergence pushes this variational distribution towards the prior distribution \( p(z) \), so that the generator can sample the latent representation \( z \) from the same distribution \( p(z) \). In this work, we choose \( p(z) \) a multivariate Gaussian distribution with zero-mean and unit variance.

### 5.7.2. Mesh Manipulation and Mutation

Users are able to manipulate the mesh of objects and then receive iterative design suggestions from Calliope. Once the user has completed their mesh manipulation, the resulting mesh is voxelated by Calliope, and the object’s associated \( z \) vector and corresponding label is used as input to conditionally sample from the network. Once the original mesh has been reconstructed from the generator, we then perform gradient ascent to reconstruct changes made to the mesh by the user. This is done by freezing the weights of the GAN and optimizing for the voxel matrix representing the designers manipulated mesh. Using the initial \( z \) vector and label, Calliope iteratively samples from the nearest neighbors in the latent space and compares these to the user-altered mesh using Structural Similarity Index Measures (SSIM) and Mean Square Error (MSE), performing gradient descent to minimize these loss metrics.

We progressively alter the step size according to the SSIM and MSE to determine how big the next guess should be. By doing this, Calliope is able to identify a vector within the latent space that best represents the designer’s manipulated mesh. This new object which approximates the user-manipulated mesh can then be encoded using our encoder architecture, and Calliope can then sample potential neighboring solutions as design suggestions to the user. Neighboring solutions are adaptively sampled, starting with random selections of neighboring objects. Our adaptive sampling approach is a base prototype and improving this approach remains a topic for future work. Once the \( z \) vector of potential design candidates has been located, the resulting voxel object is returned to the designer inside the virtual environment using
5.7 Implementation

the Marching Cubes approach, combined with a series of mesh cleaning and repairing functionalities to ensure watertightness. This process not only allows Calliope to perform mutation on a user-manipulated object, but also automates several aspects of the sculpting process by cleaning the designers manipulated mesh, thus mitigating the time required for careful mesh manipulation by the user and enabling a more rapid sculpting process.

Mutation is performed by locating the two $Z$ vectors of the two input objects within the latent space and fixing a Euclidian-distance line between the two. Calliope then samples the $z$ vectors of 2 objects evenly spaced along this line between the two input object $z$-vectors, renders the meshes of these 2 objects, and returns these objects to the user within the virtual environment. Once returned to the user, a new room is spawned containing these objects to be inspected by the user for desired features. In order to allow the user to interpolate between directly manipulated objects and GAN generated objects, we devised a technique for locating a given object geometry within the latent space. Using the method of [88], we froze the weights of the network, vectorized the manipulated mesh, and optimized a latent $z$ vector which best represented the features of the manipulated mesh within the latent space. Since multiple candidates may meet this requirement with similar probability, 4 results are sampled and displayed to the user. While this technique is present in many GAN-based approaches, it has not been implemented for a 3D voxel-based GAN architecture such as ours. This technique is crucial to our interaction pipeline because it enables a closed-loop interaction between the user and machine.

5.7.3. Computational Idle Time Costs

Due to the heavy computational cost of performing generative design coupled with direct mesh editing within VR, it is necessary to examine performance demands of multiplexing these tasks. To examine this, we stress-tested our system. It is evident
from this that mesh interpretation is the most computationally expensive operation, followed by mutation, and finally generation. While none of these tasks incur a detrimentally long wait time, they do provide a substantial period during which the user could perform other design tasks. Wait time increases rapidly with the addition of each simultaneous parallel computation process, especially interpretation of mesh editing. We should note that after 4 parallel operations, performing direct mesh-editing incurred significant performance latency, which could cause motion sickness. However, performing 4 tasks necessary to request 4 parallel computations from Calliope before the first requested task has completed is difficult. Thus, it seems unlikely that this performance issue is critical to the use of the system. These results are, of course, hardware dependent and should be explored further to better understand the computation demands of performing multiple optimization tasks in parallel with one another and a complex VR authoring system. Details of this analysis and evaluation can be found in Appendix A.

Section 5.8

DESIGN RECOMMENDATIONS

In this section, we outline principal lessons learned through the process of designing and implementing Calliope. These results are not exhaustive but are important considerations for developing future co-creative AI systems. We also compare VR vs non-VR co-creative systems where appropriate and discuss the benefits and weakness of an embodied approach.

5.8.1. Finding the Goldilocks Sample Number

As mentioned in section 5.2.3, participants expressed a need to see a multitude of options in order to inspire creativity, but not so many that they are overwhelmed.
Generative design finds as many solutions as possible that fulfill the input parameters, and subsequently produce an overwhelming number of options. Constraining this solution space is key for constructing an authoring system that is useful and beneficial for target end-users. Given the spatial nature of VR, user’s cognitive load is affected differently than 2D GUIs, and thus must be constrained appropriately. Over-constraining this sampling, however, would defeat the purpose of using such a system for inspiration and creativity. In this work, we took a greedy approach, and displayed as many samples as we could to our user without over-burdening the system, causing severe performance lag and thus inducing simulation sickness (See Section 5.7.4). Striking a balance between number of samples to present to the user is a significant challenge that hinges on constraints afforded by hardware and user bandwidth.

5.8.2. Ethical Data Use

It’s worth noting that the datasets used for training our models were either public academic datasets or proprietary datasets collected by our host institution. One method for improving the expressive capabilities of these models is to collect more data with a wider swath of data labels. However, this process has to be conducted ethically since we want to aid and support human creativity without exploiting the hard work of human artists and designers. It is for this reason that data must only be collected with the express consent of the artist or taken from the public domain. Similarly, data has to be carefully labeled in supervised tasks to ensure that bias is mitigated, and proper safeguards can be put in place to ensure the resulting model is safe for dissemination and wider use.
5.8 Design Considerations

5.8.3. Democratizing Design

Interfaces to generative design which employ intuitive metaphors for creation could potentially lower the barrier for interacting with these powerful design algorithms. Our system used a radial-data visualization inspired environment in order to provide a structural guide for a user to iteratively develop design solutions. Given the spatial affordances of VR, an embodied approach could better service a wider variety of users by enabling an intuitive exploration of a virtual space versus a 2D visual representation [89]. The benefit of this metaphor-based approach could provide an accessible platform for novice users of generative design and democratize the design process as a whole by providing an intuitive interface for novice users to sample from generative solution spaces, and guide them through the iterative design process by encouraging gradual solution space constraints.

5.8.4. Embracing Unpredictability

Often creativity support tools are employed and evaluated for their reliability of producing useful designs. However, the benefit of co-creative AI is the unexpected solutions produced by the system. In this way, systems that wish to encourage collaborative machine creativity should embrace the unpredictable nature of these systems as a source of inspiration and unexpected ideas. Similar results have been previously demonstrated by systems such as FoldIt, which leveraged a co-creative approach to demonstrate the benefit of unpredictability in these systems for inspiring and engaging users [145] [200]. In a virtual environment, however, the displayed results are more easily inspected visually, but may be more difficult to group or explore based on parametric characteristics evident in 2D GUI interfaces for generative design [110] [152]. Thus, Calliope better affords more granular visual inspection of individual generative results, relying on this inspection to curate and produce additional artifacts, whereas traditional 2D interfaces better afford exploration of the solution space through para-
metric manipulation. Investigating VR as a modality for parametric exploration of a solution space remains a fruitful topic for further inquiry.

5.8.5. Guided Sampling of Infinity

The power of generative design lies in its ability to produce ample solutions within a design space. Interfaces to such systems must take into account that this large space of potential solutions can be extremely daunting and overwhelming for users, especially when these solutions are 3D and being observed in VR. Therefore, interfaces to such algorithms must account for this by gently guiding the user through a process of iterative constraint and need-finding. In our approach, the user moved from room to room, specifying and constraining the solution space with each iteration, then viewing their results by lowering the walls. This spatial affordance is a key benefit of VR that could be leveraged beyond design of 3D objects to carefully guide a user through the iterative design process, only viewing all possible design candidates when the user desires. This micro and macro perspectives of design candidates is a key benefit of these authoring platforms in VR.

Section 5.9

FUTURE WORK AND CONCLUSION

While this initial work is promising for using the medium of VR as an interface for generative design collaboration, many opportunities exist for further investigation. While our system was able to generate and mutate between 4 different classes, additional object classes can be incorporated given adequate training of the network on a sufficient corpus of voxel object data. Creating a large latent space of such objects in the spirit of BigGAN that adheres to the data collection ethics discussed above, preserving the dignity and data autonomy of human artists would be the fruitful
subject of future work [26]. While our work focused on virtual reality, future work could expand this embodied data synthesis and exploration paradigm to augmented reality by allowing users to generate and manipulate objects in their physical environment. In this way, the spatial nature of VR which we leverage for engagement could easily be transferred to the user’s current physical environment. Finally, while we focused on interfacing to generative adversarial networks in this paper, future work could extend using VR as an interface to other methods of generative design such as topological optimization. Adapting the affordances of VR to more viscerally direct the optimization process for generative algorithms is a promising avenue for further examination. Virtual reality presents a promising platform for 3D design authoring tasks in close collaboration with generative algorithms. This initial exploratory work examined interaction possibilities of using generative adversarial networks as an active collaborator in the design process. Users are able to generate, sculpt, and delete objects, as well as mutate objects with others created during the design session. The spatial nature of VR allows the user to remain engaged in the design process exploring and designing other objects during the idle time incurred by generative computation.
Chapter 6

Multimodal Direct Manipulation in Telepresence Systems: Challenges and Opportunities

Tools supporting immersive live video conferencing (VC) have gained popularity recently across diverse application domains. A core component of the experience is augmenting video communication with multimodal interactive media. While many direct-manipulation techniques for VC communication have been proposed in existing literature, the usability and preferences for these techniques have never been formally studied. In this paper, we examine how embodied interaction democratizes content authoring, and propose a rehearsal-to-performance (RtP) framework along with a VC system, CLIO, that enables performers to directly interact with their media using voice, gesture, and external devices such as tablets. We evaluate existing operation-to-modality mappings for VC communication, as well as describe novel mappings not present in the literature. A series of studies demonstrate modality preferences and potentials for incorporating real-time direct-manipulation tools to create expressive augmented VC performances.
6.1 Introduction Multimodal Telepresence

Figure 6.1: An overview of Clio’s presentation capabilities for real-time performances, presentations, and storytelling. (a) The presenter creating behavior mappings between input interaction methods (right) and output behaviors (left). (b) A menu containing images and other virtual objects. (c) The presenter selects an image using a mid-air gesture and drags it across the screen. (d) To make an image larger, the presenter uses a two-handed pan/zoom gesture. (e) The presenter uses voice commands and mid-air gestures to display text labels on the screen and draw notes on the image.

Section 6.1

Introduction

Real-time online video communication is becoming more popular in a wide range of domains, including education [43], coding [67], creativity and art making [75], video games [148], and economics [75]. This is especially prevalent following the outbreak of COVID-19, when many daily social, professional, and educational activities moved onto VC platforms and have not yet returned to the physical world [43]. Given that a wide variety of social activities that require visual communication now take place in real-time remote VCs, visual effects for augmenting live presentations have become increasingly important to support effective communication. However, there are few tools that let people easily create and interact with visual effects in VCs, and little is known about how they are used by target end users in real-world scenarios. Unlike previous research efforts, this work focuses on formally studying the design of multi-modal direct manipulation systems to inform future interfaces.

Since this topic is significantly broad and rich, we first needed to understand how
people present visual material using current VC systems, then explore how people
would prefer to directly interact with visual media during VC presentations, and
build a system supporting these proposed interactions. The developed system can
then be used to understand how multimodal direct manipulation affects and en-
riches VC communication. To attenuate this methodology, we conducted an initial
formative design study to understand real-world use cases and limitations of existing
tools and found a tension between performer expressiveness and media augmentation.
Commercial augmented VC systems like mmhmm.ap and Cameo allow presenters to
insert graphics into their live video stream, or superimpose their video stream upon
backgrounds, but do not allow the two worlds to interact [171]. Performers are lim-
ited in their expressive capabilities due to the lack of support for direct interaction
with rich graphics, visual media, text, and drawing on the screen. This initial probe
revealed that we needed to better understand how performers would want to directly
manipulate on-screen media during live presentations.

Through a series of formative surveys, semi-structured interviews, and pantomime
studies we found that the participants employed different modes of voice, gesture,
mouse, keyboard, and tablet input to perform and control the pantomimied visual
effects essential to their use cases. However, presenters expressed concern about being
unable to imagine what their effects would look like during a real-time presentation.
We also found that enabling direct manipulation usually required technical skills and
programming knowledge that many novice users found intimidating or difficult to
acquire. Programming direct manipulation to enrich presentations is a substantial
burden in simple real-world scenarios like showing vacation photos to loved ones, Q&A
sessions during remote presentations, and other use cases where performers might be
unable or unmotivated to prepare extensively.

From our observations, we synthesized a research-to-presentation (RtP) workflow
for authoring direct manipulations using an immersive approach, and instantiated a system to support this workflow called Clio, a proof-of-concept VC system for augmenting live performances with voice, body and device driven direct manipulation. Clio enables users to easily design and integrate real-time multi-modal visual manipulations without explicit programming by offering a collection of predefined modular primitive operations derived from our formative studies. The effects resemble those that one could do in post-production using applications like Adobe After Effects, or through newscaster and weather-reporting support systems that let an external third party control the visuals using a Wizard-of-Oz approach [176]. Our workflow shifts the focus of presentation authoring from content (e.g. slide) authoring to performance authoring, which encourages presenters to focus on preparing the talk itself instead of on the artifact of the talk.

We demonstrate in a series of evaluations with live performers, audiences, and external observers, that letting performers directly interact with media contained in their presentations greatly enriches the communicative and expressive capability of VCs and encourages nonlinear presentation styles. Liberating presentations from linear constraints opens an exciting design space of nonlinear, extemporaneous storytelling and expands the application domain of VCs. In addition, the RtP immersive authoring approach built rapport between performers and machine, even when performers incorporated unfamiliar interaction techniques and machine learning tools such as voice and gesture recognition.

Our contributions of this work include: (1) Analysis of formative studies identifying diverse use cases for media augmented VCs. (2) A rehearsal-to-performance workflow and tool, Clio, for augmenting live VCs with direct manipulation using voice, body, and devices using an immersive authoring approach. (3) A series of user studies with Clio, resulting in insights into how presenters prepare and present live
6.2 Related Work

Multimodal Telepresence

VC performances augmented with interactive direct-manipulation visual effects.

Section 6.2

Related Work

Our work builds upon existing literature in several interdisciplinary domains including immersive authoring tools, dynamic media and performance interfaces, and systems enabling multi-modal direct manipulation.

6.2.1. Immersive Authoring Tools

Immersive authoring allows users to experience and verify immersive content first-hand, creating it through natural and direct interaction within the same environment [99] [128] [180]. The benefits of an immersive authoring approach is that it provides as much agency and control over system behavior to the user as is possible [216]. Immersive authoring systems for virtual environments have been widely explored in HCI research and usually involve two steps: designing virtual behaviors and content, then mapping interactions between virtual contents and users [129]. While creating virtual content usually involves programming in an environment separate from where the user experiences the behavior, immersive authoring environments blend the authoring and behavior environments [129] [255] [262] [239] [132].

SceneCtrl [261] and Window Shaping [98], for example, enable authors to create in-situ virtual scene assets and static 3D models. Similarly, Calliope supports 3D design idea generation in VR by facilitating communication between users and a creative AI through traditional sculpting techniques [239]. These ideas are extended in other projects [10] [31] [30] [260] to allow virtual contents to be animated in-situ. Visual programming has also been explored as a candidate for democratizing authorship of interactive applications [65] [92] [249] [263] [129] [177] [205]. FlowMatic, for example,
6.2 Related Work

Multimodal Telepresence allowed a user to build and test virtual interaction models in real time by connecting user input to object parameters [263]. Previous systems were limited, however, by the range of input modalities, often requiring fiduciary markers [113] [129] [188] [206], or operating only using spatial location [87] [177]. Some prior works encourage the use of midair gesture and hand pose recognition as part of the authoring process [255] [214] [248] or the use of voice [70] [189] [116] and tablet [115] [227]. While many of the above approaches use head-mounted displays and explore avatar-based virtual reality immersive authoring, in this work, we explore how VC systems provide a promising alternative venue for immersive authoring. Furthermore, a mixed modality approach that allows authors to define their behavior mappings using a variety of input methods is an unexplored problem that we investigate in this paper.

6.2.2. Dynamic Media and Performance Interfaces

The proliferation of digital technologies made interactive media an increasingly prevalent, expressive, and powerful medium for communication, art, and design [118]. An integral component of the experience is the colocation of user with their interactive media [240]. Traditional methods for creating rich and dynamic performance-driven graphics either require significant post-processing expertise, specialized preprocessing workflows, or programming. However, postprocessing is obviously impossible for real-time performances, and is only appropriate for those who have the time and skill to do complex video editing and compositing. Numerous frameworks for programming dynamic media exist like Processing [195], openFrameworks [183], D3 [45], Flash, Unity [237], and ARKit [7]. However, beside requiring significant expertise, they are limited in their support for interactive capabilities. HCI researchers have explored approaches for democratizing the creation of dynamic interactive media by prototyping novel interfaces [125] [131], sketch-based interfaces [8], direct-manipulation interfaces [133] [239], and storytelling through data [127]. Other work explores ap-
6.2 Related Work

Applications of explanatory illustrations [265][256][109] and creating mappings between user-triggered actions and animated effects [257]. Mapping human motion to digital objects [38] and digital characters [53] has also been explored in performance-based systems. Other interfaces such as SketchStudio and Kitty support user-defined relationships and events by directly manipulating elements of an illustration representing an underlying relationship graph [115, 108].

Research and commercial systems which allow manipulation of colocated media (e.g. mmhmm) are most related to our work [201] [171]; it enables users to produce real-time full-body human performances augmented with videos. We extend these ideas to encompass a broader domain of interaction methods, including voice, tablet, midair gesture, and whole body pose estimation. Furthermore, our work examines how enabling users to combine these interaction methods in any way they choose empowers them to create powerful and expressive augmented performances. Finally, our work employs an immersive authoring paradigm allowing presenters to quickly evaluate the visual effects of their authored interactive media.

6.2.3. Multimodal Direct Manipulation

Systems supporting direct manipulation of content are in common-use but extremely limit performer agency and expression. For example, newscaster and weather-reporting software create the appearance of the performer interacting with their colocated media, but the visual effect is implemented using a Wizard-of-Oz method, meaning all visual effects are controlled by an off-screen person, removing any agency or control from the performer[176]. Early real-time systems that focused on manipulating graphical elements use gestures to communicate to an audience [13]. ChalkTalk [191] and performance-driven tools [201] require users to design graphic assets and other media, then map this media to interactive behaviors. Similarly, GestuAR [248] enables creating custom midair gestures that can be mapped onto behaviors using a
head-mounted display. Other works have explored incorporating interactive digital whiteboards as part of the presentation environments [93], or incorporated wearable to assist with communication [47].

Prior work [214][53][38] demonstrates that voice commands, tablet interactions, and body movement corresponding to the presentation topic are an integral part of an effective performance, greatly enhancing the audience’s understanding of the performer’s content. Some research interfaces [117] [155] [241] have explored the potential for supporting improvised presentations and social networking apps with video filters make it simpler to generate real-time effects, but their expressiveness, applicability, and possible use cases are confined by the limited number interaction methods they support [195]. While previous research has investigated integrating direct manipulation using body or voice, these works focus on a single method of manipulation and do not support performer authorship and customizability [239] [191] [263]. However, prior literature has also demonstrated that systems that enable more than a single mode of input are more flexible for users to adopt [216], increase a user’s sense of agency and fluidity within the system [28], as well as stimulate performer creativity and audience engagement [155].

Unlike the above systems, Clio enables presenters to author dynamic media behaviors through an RtP workflow using multiple interaction methods including speech, midair gesture, tablet and others. No existing work has explored the direct manipulation of visual effects using multiple modes of input customized by the presenter for real-time video communication. In addition, an increasing number of prototype research systems exploring direct manipulation in VCS are emerging, yet these systems have never been formally studied to understand the benefits, challenges and limits to their expressive capabilities.
Section 6.3

Formative Design study

We conducted a formative design study with 8 participants to better understand the use cases and mental models of presenters when preparing and presenting with an immersive VC tool supporting content collocation such as mmhmm or Cameo. We first conducted a semi-structured interview with participants, after which they were asked to select a prepared packet of visual media from 16 topics. Using their visual media packet, they were then asked to prepare and present a presentation using a speaker and content collocation tool. Media was prepared for the participants to reduce the workload for presenters and normalize the conditions across participants. We asked participants to think-aloud during the presentation preparation process, and asked them questions regarding specific choices and preferences while they prepared. We conducted an exit interview after participants presented. 4 participants self-identified as women and 4 as men. All participants were deeply familiar with slide tools for VC presentations and also familiar with at least one immersive speaker/content collocation tool like mmhmm or Cameo.

6.3.1. Results and Discussion

Iteration: We found that, unlike conventional slide-based tool authoring that requires the presenter to prepare all visual media before presenting, participants continually added media as the presentation progressed. Preparation time was used more to select the appropriate media they may need and create an ordering. Participants playfully experimented with switching backgrounds, incorporating media, and using different system features such as the laser pointer. Playfulness and improvisation in presentation style was consistent across the preparation and performance portion of the study, where additional media was added on an as-needed basis during the pre-
sentation. Participants noted that ordering their media prior to presentation acted as a presentation outline, and not a completely concrete formulation of the presentation. Tensions emerged from the inability to display more than one image or piece of text, as well as the limited manipulation afforded by the system. “I kept blocking the stuff in my slide with my head. It would have been more useful to move the text or other parts of my slide to other parts of the screen” (P4). We probed into this, and found participants conceptualized the text and visual media in their presentations as individual elements that might coexist in a slide presentation. Manipulating and revealing these elements on command is a limitation in flexibility in the current system.

Barriers Between Worlds: Presenters experimented with various background images native to the system and those contained in the media packets until the limits of presenter and environment interaction was reached, revealing an invisible barrier between the two worlds. When backgrounds were physical places and not static images or color, they were regarded as spaces instead of objects. P2, for example, used an image from the media packet of San Francisco as a virtual background, maneuvering around and interacting with the background image as if it were a physical place in which they were immersed. Similarly, P1 used a virtual background of a coffee shop native to the system, and placed media from their packet on the tables and walls as if they decorated the space. While speaker and content colocation tools allow for the insertion of live video into media content and vice-versa, it does not allow the two worlds to interact. “The background changes are really fun...I wish I could move [the objects in the background] around like I was really there” (P4). The inability of participants to interact with their media indicates a gap in immersion when using these systems.

Modalities: Some participants attempted to use different modes of interaction
while exploring the system during the presentation preparation phase of the study. “I don’t see why I can’t get a robot to change the slide for me when I say ‘next’” (P3). The tension between the presenter and media worlds was also evident in pain points around tool usage in current systems. “I kept getting confused. I thought [the laser pointing feature] was for drawing and was confused when my pen marks wouldn’t stay on the screen. Would love to draw on the screen using this tool and my iPad” (P3). Some expressed a desire to use different modalities in conjunction with each other in order to manipulate their content. “It could be like Star Trek where we tell something to ‘zoom and enhance’ and it enlarges [the image] automatically” (P1). Others noted that constraining the mode of manipulation to the mouse posed key limitations. Participants also remarked that reliance on using the mouse as a primary modality of control were limited. “You couldn’t use the cool stuff on your phone. You’d have to have tiny fingers to drag things” (P1). Employing multimodal direct manipulation while maintaining an immersive approach could ease these tension, blending the world of the presenter and their media together and approaching the rich interactive potential evident in augmented and virtual reality.

### Section 6.4

**Formative Pantomime Study**

Understanding how to blend performer and media content is difficult for a variety of reasons, including the wide variety of different embodied interaction possible (e.g. voice, body pose, etc.) and the lack of familiarity that many potential users have with these machine-learning enabled tools. Furthermore, controlling a system that uses a mixture of input modalities presents challenges because it is unclear which mode of interaction is appropriate for controlling the authoring system. To better understand the need of presenting content in immersive environments using multimodal direct
manipulation, we deployed a questionnaire and two-phase semi-structured interview with pantomime study of potential user presentation processes using a think-aloud methodology. The study took place in two 30 minute sessions with 8 participants. 4 of our participants self-identified as women and 4 as men. At the conclusion of the first phase, participants were asked to think of a use case where they might present something in an immersive environment using direct manipulation. The second phase took place on a separate day from the first, and participants were asked to pantomime their use case twice using whatever input modality they felt appropriate. Participants first pantomimied their use case while thinking-aloud, describing their thought processes and what they were trying to accomplish as they proceeded. During this pantomime, participants were encouraged to use any method or tool they deemed necessary to communicate their idea including physical props. After completing their pantomime the first time, participants were asked questions regarding their choices and rationale behind different choices and behaviors. Participants were then asked to perform their pantomime again uninterrupted. Based on our observations and their spoken explanations, we segmented the data into individual interactions that were then analyzed along several dimensions. This approach allowed us to compare the use of gesture, voice, external devices, mouse and keyboard, and identify common usage patterns.

6.4.1. Results

We refer to the media manipulation behaviors proposed by participants, such as selecting, arranging, highlighting, and zooming an element as operations; Table 6.2 shows how many participants suggested each operation. They also suggested a variety of input methods, such as voice, gesture, and mouse, and we call these interactions. A complete list of the use cases, interactions, and operations evident in these pantomimes can be seen in Table 6.1 with further details in Appendix B. In interviews, 75% of participants indicated that they would prefer to watch a 30 minute presenta-
Table 6.1: Participant-suggested use cases from formative Pantomime Study with accompanying operations and interactions. *Interaction* refers to the input method used by the participant during the pantomime and *Operation* refers to the behavior performed by the interaction.

<table>
<thead>
<tr>
<th>Use Case</th>
<th>Operations</th>
<th>Interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>p1 Portrait Portfolio Consultation</td>
<td>arrange, select, open/close menu, group objects, zoom 2D, highlight, laser pointer, dismiss, make transparent, expand collection</td>
<td>Gesture, Voice</td>
</tr>
<tr>
<td>p2 Project Presentation</td>
<td>arrange, select, open/close menu, dismiss, conjure, zoom 2D, laser pointer, next slide, previous slide</td>
<td>Keyboard, Mouse, Gesture</td>
</tr>
<tr>
<td>p3 Conference Q&amp;A</td>
<td>conjure, text display, annotate object, annotate air, open/close menu, select, next slide, previous slide</td>
<td>Keyboard, Mouse, Voice</td>
</tr>
<tr>
<td>p4 Interactive Demo Session</td>
<td>open/close menu, arrange, dismiss, rotate 3D, zoom 2D, zoom 3D, highlight, select, activate</td>
<td>Keyboard, Gesture</td>
</tr>
<tr>
<td>p5 End-of-year Student Presentation</td>
<td>open/close menu, highlight, dismiss, collapse collection, group objects, dismiss, pull audience content into screen, arrange, next slide, previous slide</td>
<td>Gesture, Mouse</td>
</tr>
<tr>
<td>p6 TA Session</td>
<td>arrange, conjure, zoom 2D, annotate object, highlight, rotate 3D, dismiss, tangible proxy, select, next slide, previous slide</td>
<td>Tablet, Gesture, Voice</td>
</tr>
<tr>
<td>p7 Vacation Photo Presentation</td>
<td>annotate air, display text, arrange, trigger, dismiss, select, open/close menu, conjure, annotate object, highlight, tangible proxy, create virtual copy, group objects, next slide, previous slide, poll/quiz, add shape, screen grab</td>
<td>Gesture</td>
</tr>
<tr>
<td>p8 ASL Tutorial and Conversation</td>
<td>text display, conjure, track image (to gesture), screen grab, next slide, previous slide</td>
<td>Keyboard, Gesture</td>
</tr>
</tbody>
</table>
Table 6.2: Occurrences of specific operations presented by participants during their use case in Phase 2 of the study. Full list of operations can be found in Appendix B

<table>
<thead>
<tr>
<th>Operation</th>
<th>Count</th>
<th>Operation</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group Objects</td>
<td>3 (38%)</td>
<td>Expand Object Group</td>
<td>3 (38%)</td>
</tr>
<tr>
<td>Next Item</td>
<td>6 (75%)</td>
<td>Previous Item</td>
<td>6 (75%)</td>
</tr>
<tr>
<td>Tangible Proxy</td>
<td>2 (25%)</td>
<td>Create Virtual Copy</td>
<td>1 (13%)</td>
</tr>
<tr>
<td>Screen Grab</td>
<td>1 (13%)</td>
<td>Push/Pull Content from Chat</td>
<td>1 (13%)</td>
</tr>
<tr>
<td>Poll/Quiz</td>
<td>2 (25%)</td>
<td>Laser Pointer</td>
<td>4 (50%)</td>
</tr>
<tr>
<td>Add Shape</td>
<td>2 (25%)</td>
<td>Remove Shape</td>
<td>2 (25%)</td>
</tr>
<tr>
<td>Alter Shape</td>
<td>2 (25%)</td>
<td>Composite Shape</td>
<td>2 (25%)</td>
</tr>
<tr>
<td>Rotate 3D</td>
<td>2 (25%)</td>
<td>Make Transparent</td>
<td>2 (25%)</td>
</tr>
<tr>
<td>Activate/Trigger</td>
<td>4 (50%)</td>
<td>Select</td>
<td>8 (100%)</td>
</tr>
<tr>
<td>Open Menu</td>
<td>6 (75%)</td>
<td>Close Menu</td>
<td>6 (75%)</td>
</tr>
<tr>
<td>Highlight</td>
<td>5 (63%)</td>
<td>Draw on Object</td>
<td>5 (63%)</td>
</tr>
<tr>
<td>Draw in Air</td>
<td>5 (63%)</td>
<td>Arrange Objects</td>
<td>6 (75%)</td>
</tr>
<tr>
<td>Zoom 3D</td>
<td>2 (25%)</td>
<td>Zoom 2D</td>
<td>6 (75%)</td>
</tr>
<tr>
<td>Dismiss</td>
<td>5 (63%)</td>
<td>Conjure</td>
<td>6 (75%)</td>
</tr>
</tbody>
</table>

6.4 Formative Pantomime Study

Multimodal Telepresence

...tion that used visible gestures, voice, or other noticeable interactions versus hidden interactions like mouse and keyboard or discreet static hand gestures. Similarly, 63% of participants indicated that they would prefer to perform a 30 minute presentation using visible interactions, and 38% of participants indicated that they would like to both watch and perform this way. This means that while many participants indicated they would prefer to watch presentations using dynamic embodied interactions, they might prefer to give presentations using discreet modalities such as mouse and keyboard or static gestures, and vice-versa.

6.4.2. Design Considerations

Based on the above observations, we synthesize the following design considerations and goals to guide the developing a system to support for semi-extemporaneous presentations in virtual environments. In the next section we describe a workflow extracted from our formative interviews that supports these design considerations.
6.4 Formative Pantomime Study

D1. Maintaining Immersion: To better nurture presenter intuition for how the system will support their presentation, there should be little difference between the interface in which content is prepared and the interface in which it will be presented. Embracing immersive authoring shifts the focus of the presenter from preparing slides and media to the act of delivering the presentation.

D2. Experimentation with Different Interaction Methods: Presenters should be able to experiment with various interaction methods and tools as part of their preparation process. This could alleviate the mistrust expressed by participants when using tools and interaction methods with which they are not familiar. Subsequently, a presenter should be able to rapidly switch between interaction methods to experiment with as many as possible, as well as to adapt their presentation to their current contextual needs. A system that leverages the tools afforded by virtual environments should integrate exploration and experimentation of interaction methods as part of the presentation preparation process.

D3. Nonlinear Presentation Style: As discussed above, one of the weaknesses of traditional virtual presentation tools is their strict sequential nature, making improvisation and extemporizing difficult. One of the benefits of systems that collocate the presenter with their content is the flexibility to support playfulness and improvisation. A system designed to blend the worlds of presenters and their content should support flexible modes of presentation to enable a variety of presentation styles.

D4. Direct Manipulation: Finally, blurring the worlds of the performer and their digital media requires direct manipulation of content. Supporting a variety of methods to perform such manipulations enables presenters to ensure modes of interaction are appropriate for specific presentation media, as well as present material using means most intuitive to them.
Speaker/content colocation with systems like mmhmm is one form of immersion, which we extend to incorporate multimodal direct manipulation. Based on the observations from our formative studies, we conceived a workflow that addresses the participants’ needs and describes a mental model exhibited by participants while authoring and presenting immersive speaker/author colocation performances during our formative studies. This rehearsal-to-performance model replaces traditional approaches to presentation preparation with an immersive authoring approach (D1) that supports direct manipulation of media content (D4). Our model is based on a rehearsal phase that supports presenters in iteratively preparing their presentations. The key distinction between the RtP framework and conventional workflows using slides presented within a VC to present content is that that presentation media is authored using the same interface with which it is presented. Performer behavior with the interface changes between the rehearsal and performance phases, but the interface itself does not. During the rehearsal phase, presenters experiment with different interaction methods and operations to identify which behavior mappings between operations and interaction methods are appropriate for the context of their presentation and appropriate for the media they are presenting (D2). During the rehearsal phase, presenters can organize their media and experiment with different behavior mappings using the mapping interface. Once presenters feel sufficiently prepared, they shift to a performance phase in which they use the outline as the architecture of their presentation. Our workflow shifts the focus of presentation authoring from content (e.g. slide) authoring to performance authoring, which encourages presenters to focus on preparing the talk itself instead of on the artifact of the talk. Since the presentation
6.5 Rehearsal-to-Performance Workflow  Multimodal Telepresence

media is not bound to any sequential order in this workflow, presenters can discuss their presentation material in any order they may choose (D3) and add additional media or change any element of their presentation as needed. Figure 6.2 shows this workflow.

6.5.1. Interaction Design Overview

A presenter begins by importing media, such as images, videos, external web links, and graphs, into the system. After importing the media, the presenter enters the rehearsal phase where they organize their presentation (See Figure 6.1). The default behavior mapping between operations and interaction methods lets the presenter use a mouse to interact with their imported media. Clicking the menu bar on the left side of the screen makes a tray containing thumbnails of the uploaded media appear. This tray can be moved to any side of the screen by clicking and dragging to the desired location. The user can scroll through media in the open tray and choose to click and drag entries onto the rest of the screen (D3). Here they can perform any supported operation (Section 6.6.2) using the mouse and keyboard as the controller.
If the presenter clicks the “begin mapping” button in the top right hand corner of the screen, a menu appears that lets them customize the mapping of operations to interactions (D2). After modifying the mapping, the menu closes and the presenter can freely experiment with their new mapping. This enables presenters to quickly evaluate different interaction approaches and find the behavior mapping between interactions and operations appropriate for their performance. An overview of the mapping interface can be seen in Figure 6.3.

Section 6.6

CLIO System Design

To explore the feasibility of a RtF workflow, we created CLIO, an immersive authoring and presentation support tool for developing and delivering semi-extemporaneous presentations in virtual environments. CLIO is prototyped using a proprietary research system that allows HTML, CSS, and JavaScript code to be overlaid on top of a camera video feed. Much like other VC augmentation systems (e.g. mmhmmm), CLIO can be fed directly into the video source of commercial VC systems, using the built-in camera common to contemporary commercial laptops, thus requiring minimal setup. The main interface and logic is coded in Javascript that is rendered in a Chromium browser embedded within the application interface. We refer to the supported input techniques, including voice, mouse, body pose and midair gesture control, as interactions, and the resulting behaviors, such as arranging media and drawing on screen, as operations. The user is able to create behavior mappings that associate an interaction with an operation using a behavior mapping menu.
Figure 6.3: CLIO’s behavior mapping system. (a) Presenter is able to activate mapping menu at anytime during rehearsal or presentation. (b-1) Inside the menu, the presenter selects one operation from the left side of the menu (b-2) the presenter selects one interaction method from the right side of the screen.
Figure 6.4: State-based dynamic gesture detection system. (a) Finger-joint angles used for hand pose detection. (b-1) We detect one of eighteen static hand poses. (b-2) We then look for sequences of static hand poses to detect the beginning and ending of dynamic gestures. (c-1) First state gesture in the sequence (closed fist). (c-2) Second state gesture in the sequence (extended index finger). (d) User performs dynamic gesture. (e) User ends the dynamic gesture by returning their hand from the second state pose to the first state pose.
6.6 CLIO System Design Multimodal Telepresence

6.6.1. Interactions

This section describes the interaction methods supported by our system. We chose them because of their frequent occurrence within our formative study.

**Mouse, tablet touch, and keyboard:** The mouse is the default mode of interaction with our system since users from our formative study reported the most comfort and familiarity with this interaction. We also provide similar support for tablet touch interactions, which behave similarly, using a finger or stylus. Keyboard interactions can perform operations and provide text for operations that need it.

**Voice:** Presenters can perform specific operations (e.g. displaying text) using their voice. They can specify specific keywords that will perform an operation. Two modes of performing operations are supported, inline and stand-alone. Inline mode parses all detected speech for keywords without requiring the speaker to pause. It is more discrete since the presenter can weave keywords into their presentation text and perform operations in a hidden manner. The stand-alone method requires a break in verbal speech before a keyword can be recognized and perform an operation. Switching between these two keyword modes can be done through the mapping menu during rehearsal or presentation if needed.

**Gesture:** Midair gesture and body-driven interactions are also supported by CLIO. The underlying system provides hand-joint and body-pose data based on the video feed. Our algorithm calculates poses by sampling the positions of fifteen hand-joint angles and averaging these time-series joint angle samples to retrieve the pose of each finger. From here, each finger position is classified open, closed, or bent. We define a hand gesture to be a collection of required states for each finger. If each finger approximates the required state associated with a specific gesture, that gesture is considered active. Our initial implementation of CLIO supports 18 static gestures following the example of previous studies which revealed that typical midair
gestures are comprised of two components; local properties meaning how the hands are posed, and global properties, indicating the location and movement direction of the palms [248]: closed fist, open palm, index finger extended, middle finger extended, ring finger extended, thumb extended, pinkie extended, two-finger pinch, four finger pinch, index and thumb pinch, index and middle extended, index middle and thumb extended, index and thumb extended, index and pinkie extended, cupped hand (open), cupped hand (closed), index middle and thumb bent. To enable dynamic gesture recognition, we define a dynamic gesture to be a sequence of static gestures. The static gestures must be made in a specific order to perform the operation associated with the gesture. Once performing, the operation remains active until the gesture sequence is repeated in the opposite order. For example, the initial static gesture for ‘grab’ is an open palm, the second static gestures is a closed fist. Performing these in sequence performs the operation mapped to this gesture by the user in the behavior mapping menu. To end the operation, the user shifts their gesturing hand back to the open palm that began the interaction sequence. Two-handed interactions are only recognized by CLIO when both hands are present and performing the requisite gesture sequences.

6.6.2. Operations

Operations are the ways in which a presenter can manipulate their presentation content using their chosen interactions. We support nine operations, summarized below and described fully in Appendix B (See Figure 6.5). We chose these because of their frequent occurrence within our formative study as well as their prevalence in existing literature [2].

- Arrange: enables the user to move virtual objects around the screen
- Pan/Zoom: enables participant to enlarge or shrink a virtual object
Figure 6.5: Operations Supported by CLIO. (a1) Arrange: the user places their hand over the image and creates the Arrange gesture. (a2) the user moves the image to the desired location and stops making the Arrange gesture to place the image in the desired area. (b1) Pan/Zoom: the user places both hands over the image and makes the Pan/Zoom gesture. (b2) the user moves their hands away from each other to enlarge the image. (c1) Draw: the user makes the draw gesture by raising their index finger. (c2) The user draws on the screen with their index finger. (d) Display Text: the user makes the Display Text gesture (raises two fingers) and speaks to display their spoken words as text on screen. (e1) Conjure: The user makes the Conjure gesture (e2) Images appear on screen where Conjure gesture was initiated. (f1) Dismiss: The user makes a dismiss gesture. (f2) Image disappears when user makes Dismiss gesture over visible image. (g) Highlight: similar to draw gesture, the user makes the highlight gesture (three fingers) to draw highlights on the screen text. (h1) Next/Previous: the user makes the Next or Previous gesture (thumb right and thumb left respectively) to hide all presently visible media and text and move to the “next” or “previous” image in the media library, similar to moving to the “next” or “previous” slide in a slide presentation.
• Draw: supports annotation anywhere on the screen or on a virtual object. Colors, brush size, and opacity can be changed

• Text Display: User specifies what text they would like to display on the screen and where

• Highlight: Creates a glowing effect around a virtual effect

• Conjure: Makes a specific virtual object appear on the screen

• Dismiss: Makes a specific virtual object disappear

• Next: Cycles to the next virtual object in the media tray (a more general approach to “next slide” in a traditional slide presentation)

• Previous: Reverts to the previous virtual object in the media tray (a more general approach to “previous slide” in a traditional slide presentation)

Section 6.7

Study 1: Validation of Formative Study Use Cases

To understand whether Clio effectively supports the immersive authoring and multimodal direct manipulation needs expressed by presenters in our formative study, we conducted a follow-up investigation with some of our original participants. Five participants (p1, p2, p3, p5, p8) had pantomimed presentations that were based on operations that we implemented in Clio; the other three (p4, p6, p7) relied heavily on operations like 3D object rotation and tangible proxies that we did not include in our first version. These operations were only used by a few participants, and fully understanding their needs and uses requires a subsequent independent investigation.
Excluding these therefore will not affect our findings. Each of the five invited participants used Clio to create and perform an actual presentation that closely resembled their original pantomimed presentation. We gathered statistics on their activities and conducted a post-task interview that included both Likert scale and open-ended questions. Our Likert questions used a scale of 1 to 5 with “1” signifying “not at all” and “5” indicating “very much so”.

**Results and Discussion:** All participants were able to present their use case with Clio. Presentations on average took 2.42 minutes (SD = 0.55) and required an average preparation time of 24.32 minutes (SD = 2.52). This time includes a tutorial session to familiarize participants with the system, as well as time for the participants to experiment with the various capabilities of Clio. In the following sections, we will reference the question from our study which informed our insight as well as the place where the question can be found in the appendix. Overall, participants found the system easy to use (Q:B.4.17 “It was easy preparing and giving virtual presentations using the prototype” AVG = 4.67 SD = 0.33), intuitive (Q:B.4.19 “It was intuitive to prepare and give my presentation using the prototype” AVG = 4.33 SD = 0.64), and capable of supporting the use case conceived during the formative study (Q:B.4.31 “My presentation closely resembled my pantomime during the previous phase of the study” AVG = 4.25 SD = 0.96). “It’s so fun! I can imagine ways this could be used for art as well as presentations. It’s so playful” (p1). A full list of questions asked can be seen in Appendix B.4.

Commonalities in behavior mappings and interaction preferences were evident among presentations with similar contexts and levels of formality. For example, p2 and p5 both involved presentations in professional environments with more structured presentation requirements, and subsequently used more traditional keyboard/mouse behavior mappings. P1 and P8, however, presented use cases that were more informal
and playful, and thus embraced more overt behavior mappings such as arranging using midair hand gestures and pan/zooming using the custom keyword “enlarge”. Most participants used behavior mappings similar to those presented using pantomime during the formative study. P3 also presented a formal use case (conference Q&A) and pantomimed mostly mouse/keyboard behaviors during the formative study. However, while using CLIO, p3 deviated from their original pantomime, and used hand gestures discreetly to switch presented media. When asked about this change, p3 explained that they were comfortable using hand gestures for small interactions, and that using these gestures instead of the mouse would allow them to potentially move around their physical space while presenting. While most of the use cases suggested by our participants were conceptualized as remote VC presentations, p3’s adaptation suggests that the benefits of immersive multimodal mapping could transfer to in-person presentation contexts.

Three participants (p2, p3, p5) noted that the system caused them to rethink their presentation approach. “I had to rethink my presentation a bit since the system works different than PowerPoint. You can go in order but you can also deviate and that has me rethinking my talk from the ground up” (p2). Furthermore, participants expressed curiosity about what it would be like to compose a presentation knowing the capabilities and workflow of CLIO ahead of time, as opposed to implementing a previously-conceived use case. For this purpose, we performed a second user study to examine how users prepare new presentations.
6.8 Clio Validation Study 2

Section 6.8

Study 2: Native Immersive Direct-Manipulation Presentations

To understand how immersive authoring and multimodal direct manipulation enabled by Clio affects computer-mediated communication, preparation, and presentation for both audiences and presenters, we designed a 3 participant group study methodology. These groups comprised the presenters, the audience, and an external group of commentators. Our study sought to understand how an immersive authoring approach affects presentation preparation and performance, how different modalities used in presentations are received by audiences, does an RtP workflow develop trust between user and machine, and how is the presentation of information different using direct manipulation?

6.8.1. Participants and Task

Our study comprised 6 presenters (referred to under the moniker "pr") who prepared and gave presentations, 12 audience members (referred to under the moniker "a") who watched the presentations, and 5 external commentators (referred to under the moniker "c") who observed the presenters and audience members. Of the 6 presenters, 2 identified as women and 4 as men, and ranged in age from 28–44 with an average of 32. Of the 12 audience members, 7 identified as women and 5 as men, and ranged in age from 24–61, with an average of 35. Of the 5 external commentators, 3 identified as women and 2 as men, and ranged in age from 26–62 with an average of 44. All participants had prior experience with virtual presentations; 78% indicated that this experience was largely with slideshows shared over Zoom, Google Meet, or Microsoft Teams. Additionally, all had experience giving virtual presentations with
the exception of one member of the audience group. One member of the audience had previous experience using voice and gestures to interact with virtual presentations through their art-making practice.

**Presenters.** All presenters were presented with ten potential presentation topics to choose from similar to the prepared media packets used in our formative study (See Section 3.1). After selecting their presentation topic they were given a collection of media that they could use use in a presentation on their topic. No topic was chosen more than once. Presenters were then introduced to Clio and given a brief tutorial on Clio’s use. Following this tutorial, presenters were then left alone to rehearse with Clio, changing interaction/operation mappings and experimenting with their assigned media packet. Once all presenters were adequately prepared, our virtual audience was brought in to watch the presentations. Each presenter took a turn presenting their rehearsed presentation using Clio to the live virtual audience. After all presenters had completed their presentations, they were given an exit questionnaire containing free response and Likert scale questions to probed their experience presenting with the prototype, as well as the overall rehearsal-to-performance workflow.

**Audience.** Audience members were gathered on Microsoft Teams to watch the virtual presentations. After watching each of the presentations, audience members were given an exit questionnaire with free response and Likert scale questions to examine their experience watching the virtual presentations. Both the virtual presentations and the audience gallery was recorded during the live presentations.

**Commentators.** After the live virtual presentations were completed, the recordings of the audience gallery and presentations were shown to 5 commentators. We asked these commentators to make observations about the audience and overall impression
of the live presentations using a questionnaire.

**Section 6.9 Results**

All 6 presenters were able to successfully complete the rehearsal and performance sessions. On average, participants required 31.06 (SD=1.45) minutes to complete the tutorial and rehearsal process for their presentations, which ran 2.52 (SD=.34) minutes on average. Much of this time was spent familiarizing the presenter with the system, as well as allowing the presenter to explore and experiment with as many mappings as they wanted.

**6.9.1. General Impressions and Usability**

Presenters indicated that they liked giving their presentation using Clio (AVG=4.67 SD=0.49) and that they would like to experiment giving their talks using different interaction mappings. “I feel like I played it safe and really would have wanted to use voice or gesture or something more flashy... I think it would be better at keeping people’s attention” (pr5). All presenters indicated that they would rather watch a presentation using Clio than traditional presentation tools (AVG=5.00 SD=0.0). Presenters overall agreed that Clio was easy to use (AVG=4.33, SD=0.67), intuitive (AVG=4.67, SD=0.33), and fun (AVG=5.0, SD=0.0). Audience members were generally excited by the promise of Clio as a presentation system. “I was intrigued by the use of different presenting tools... It was interesting to see as an audience part of what the presenter was seeing—almost felt like an interesting two-way mirror [where we were] ‘in the room’ with the presenter” (a3). Overall, the audience indicated that they enjoyed watching presentations using Clio (AVG=4.75, SD=0.25), felt generally engaged (Avg=4.66, SD=0.47), and would rather watch a virtual presentation
6.9 Results

Multimodal Telepresence

Figure 6.6: Average Likert Results from Validation Study for Presenters (PR) and Audiences (A) on a 5 point scale where 5 indicates “very much so” and 1 indicates “not at all”. Axis labels reflect components of Clio. Color and size of bar indicates average Likert score reported for each dimension evaluated (See Legend)

using Clio than using traditional presentation tools (AVG=4.66, SD=0.33). While audience members indicated that they had noticed gesture and voice interactions, none had noticed the mouse or tablet being used. Preference was given to inline speech interactions vs stand-alone interactions due to the pause required for the latter to function properly being “distracting” (a12). Similar to remarks made by the presenters, audiences indicated that they felt more engaged because the presentation had a more conversational feel, and thus they were more inclined to ask questions and interact with the speaker (AVG=4.33, SD=0.96). “It feels like it’s more natural to have moments where questions can come up organically and the presenter and audience don’t feel like they’re interrupting each other” (a2). Commentators agreed...
that the audience was engaged (AVG=4.75, SD=0.2) and the presentations were more engaging than traditional presentations (AVG=5.00, SD=0.0). As the presentations progressed, more virtual items were moved onto the screen by the presenter. A few times during the presentations, the images on the screen occluded the presenter’s face from the audience, or images consumed the majority of the screen. “[the audience] start looking off. Checking their phones. They lose interest when there’s too many things on the screen” (c5). Commentators also noted that the gesture interaction seemed to garner the most attention, followed by voice. However, commentators observed that the engagement produced by voice seemed to wane over the course of the presentations. “The excitement wears off... it feels like a natural part of the presentation.” (c5). Similar to the audience response above, commentators also did not notice the mouse or tablet being used. Additionally, speech interactions using the inline method discussed in section 5.1 also went unnoticed. When asked about this commentators responded “Sometimes I could follow the person’s hands and I knew what was about to happen but sometimes things just moved and I had no idea how... it seemed like magic” (c1). Full results of Likert study can be seen in Figure 6.6.

**Modality Preferences** All interaction methods and operations were used at least once across the 6 presenters and all participants used at least one gesture and one voice interaction. Participants reported that they experimented with at least 3 interaction methods during their rehearsal period, and ultimately gravitated towards using one more predominantly than the others. All presenters experimented with gesture and mouse, while half experimented with speech. 50% of presenters reported that their overall favorite interaction method was midair gestures, 33% favored speech, and 17% preferred the mouse. Similar to Study 1, we observed a correlation between content formality, presentation style, and behavior mappings. For example, p3 presented “architecture” and employed a more formal style and conventional behavior mappings
6.9 Results Multimodal Telepresence

using mouse and keyboard. Compare this to p2 who presented “weird trees”, employing mostly voice and overt gesture behaviors. Individual preferences for specific interaction methods varied based on presenter choice of mappings. “Controlling the presentation with speech and gestures made the presentation feel more ‘in person’ than the zoom presentation + head screen” (pr2). While presenters on average suggested they would rather deliver a virtual presentation using the prototype than with traditional presentation tools (AVG=4.0, SD=0.49 ) they also indicated that different mappings were appropriate for different presentation contexts. “For something that has to be tightly scripted (time limit, etc.) doing prerecorded video or doing a slide deck with built-in animations would probably be preferable to having the presenter need to move things around” (pr1).

6.9.2. Comparison to Existing Tools

Here we outline presenter and audience feedback comparing existing commercial tools with Clio. Participants were asked before and after presentations about their familiarity and experiences using various platforms. Since all participants (audience, presenters, and commentators) had significant familiarity with one or more commercial VC tool, participants were asked to directly compare their previous experience with these tools to Clio (See Appendix B.1 q6–25, B.2 q23,27, B.3 q3, 28, B.4 q3, 28, B.5 q5–11).

Slideshow w/VC. All audience and presenters were familiar with traditional VC tools as well as slideshow presentation software used in combination to present graphics and media content. Several key pain points and distinctions were outlined by participants. “Changing the slides can be clumsy [I have] to say “next slide” on zoom” (a10). The participant references a context where the person presenting in a VC session is not the same person controlling the slides, or does not have their permissions
correctly configured to present their material, and thus must rely on someone else to do so. Since CLIO feeds directly into the video stream of commercial VC’s, it is possible for CLIO to alleviate this pain point by enabling the presenter to change visual media using gestures or other non-vocal modalities similar to how they would change slides. Similar sentiments that sharing the screen and changing slides is clunky was a common theme among both audience and presenters (a1–3, a5–11, pr1–4, pr6). These participants noted that using voice and gestures allowed them to change media content without touching their computer, potentially alleviating transition clunkiness, and could eliminate needing conference hosts to change slides for them. Additionally, presenters pr2–4, and pr6 noted that this could allow them to move about the room and still control the media displayed on the screen, untethered to their computers. Others (a1–2, a4–9, a12, pr1–2, pr5) also expressed that slideshows often made it feel as if there “was no energy in the room” (a4) or that there was “some kind of disconnect between the person talking and everyone else” (a9). Participants noted that colocating speakers with presentation content increased “feeling the presence of the speaker in the meeting room. . . I felt like I couldn’t check my email or twitter” (a2). This result corroborates existing literature on the attention effect of inserting live video feeds into slideshow content [43], demonstrating that live video feeds support increased attention when colocated with presentation media.

**Speaker and Content Colocation.** Those familiar (a1, a3–5, a9–10, pr1, pr3, pr5–6) with VC platforms that allowed live video footage of the speaker to be inserted into the presentation (e.g. Cameo, mmhmm, etc.) expressed the limitations of engagement resulting from colocation of presenter with content, noting that “it feels static. . . [there is] some kind of disconnect between the person talking and the rest of the stuff on the screen” (a1). Customizable manipulation using voice and gesture mitigates the gap between presenter and content by providing a bodily connection
between performer and media behavior. “Controlling the images and text with speech and gestures made the presentation feel more ‘in person’ that the zoom presentation plus head screen. It was very interactive.” (pr2). All presenters echoed this sentiment, noting that manipulating with their media using different techniques created a stronger sense of agency over their presentation. “My past virtual presentations were not as dynamic. This type of virtual presentation allowed for a lot of movement and left the door open to improvisation. With audience feedback you could be more fluid with what material you wanted to highlight and what material you could leave to the side.” (pr3). Colocation of presenters with their media resulted in a more playful environment, enabling a looser presentation style that could encourage improvisation. Audience members (a1, a3–5, a9–10) highlighted an increased sense of presenter “presence in the room...[the presenter] moved stuff around the screen like they could move [physical] things around in a real room” (a10). This relationship between agency over digital objects and the resulting verisimilitude of virtual reality environments is evident in prior literature [76, 99, 180]. While some research suggests that multimodal manipulation of objects enables users to interact with media in the manner most suited to their needs and abilities, little work has documented the impact multimodal manipulation has on the believability of a VC environment.

Nonlinear and Brainstorming. (2 presenters and 4 audience members) Some presenters and audience members (pr1, pr3, a3–5, a12) were familiar with nonlinear brainstorming tools such as Miro and Whiteboard, and noted how often brainstorming occurred during VC discussion while simultaneously using these tools. Some (a4, a5) suggested that during such brainstorming co-design sessions, Clio could be “Easier to be more responsive to [the] audience without disrupting flow” (a4) and thus could support brainstorming and nonlinear tasks where a single person can facilitate the session. Others (a3, a12) were more interested in the multiuser aspect of these
nonlinear systems, and suggested that developing methods for allowing multiuser interaction (e.g. Miro) would be an exciting direction for the CLIO, and necessary for supporting brainstorming, mindmapping, and other nonlinear tasks. Both presenters, however, envisioned using CLIO alongside existing nonlinear systems: “my dream is to use this with Miro (which I use a lot at work) to teleport me to different parts of the board by voice and use gestures to navigate rather than mouse when presenting or facilitating a workshop” (pr1).

**Section 6.10**

Discussion

By employing multiple modes of direct manipulation, presenters blurred the barrier between themselves and their content and transformed static images into interactive characters and responsive environments. Some presenters used voice interaction to summon, move, and dismiss specific images from their packets, which they named and interacted with as if they were characters. Gesture interactions were also used to move and interact with these images, creating small, playful scenes between presenter and characters. “I loved the dogs! They each had a unique personality and relationship with [the presenter]. Reminds me of playing with my dogs at home” (a2). All audience members regarded characters animated using CLIO as participants in the scene. Perceiving static images as characters interacting with the presenter suggests that the barrier between performer and content is sufficiently blurred. Other presenters held two images at a time, and made them speak to each other as if they were hand puppets. One presenter enlarged an image of a tree, and drew plans for an imaginary tree house to be built on top. “I kept forgetting that [the presenters] weren’t actually holding [physical objects] and that they were just moving images around the screen with their hands. It fooled me!” (a4). These performances suggest that not only
employing bodily direct manipulation blends the gap between presenter and content, but maintains an immersion seamless enough to author interactive skits with digital images.

Audience members remarked that using gestures resulted in more motion being evident during the presentation and this helped keep their attention. “It was different [than traditional virtual presentations] because you weren’t looking at a still image with text. There’s a person there with you moving things around so you pay attention. It’s cool.” (a4). Presenters indicated that they were skeptical of using interaction techniques that they were unfamiliar with before rehearsing (AVG=4.33, SD=0.33), but that they were much more trusting of these interaction methods afterwards (AVG=4.33, SD=0.67). Part of the reason was that participants felt a greater rapport with the system due to the rehearsal process (AVG=4.67, SD=0.47) and being able to quickly view the interaction mappings they had created during rehearsal (AVG=4.67, SD=0.49). “I thought that this was pretty vital. Knowing where to stand and what distances were relevant were important. I think in general it’s valuable to make practice as close as possible to performance.” (pr2).

**User/Data Colocation and Screen Occupancy.** Audience members suggested that they were more engaged with the virtual presentation using Clio, partially because they could see the speaker better than in traditional virtual presentations (AVG=4.4 SD=0.64) “It made the presentation more personalized and allowed me to follow what they were saying. Since they could gesture to the images rather than the images occupying the whole screen meant that I could better associate what they were saying with what I was seeing. Versus having to just use my brain to read words on a screen and not being able to listen to the speaker simultaneously” (a9). Presenters reported that their attention felt ‘split’ (pr6) as more images, text, and annotation were added to the screen. “I could really only manage three images on
the screen at a time. More than that I started feeling overwhelmed and had trouble grabbing a specific image [using gestures].” (pr2). Interestingly, this sentiment was echoed by presenters who primarily used the mouse as opposed to gestures to manipulate objects on the screen. This suggests that interaction method and mapping are not the main factor contributing to the burden caused by screen occupancy. Despite this, all participants still felt more “connected” (pr5) to their media due to being colocated with it. “I could indicated specific parts of the image using my body, or even reference common items between two images simultaneously. It was easy and really neat.” (pr3).

**Improvisation vs Linear Planning.** After the rehearsal, all presenters remarked how they felt Clio lent itself to a more improvised mode of speaking. “It felt less formal, like I was rehearsing how I wanted to walk through my content and ideas and less about structuring presentation materials. I was improvising with support from the software” (pr1). This sentiment carried through the rehearsal into the performance as well. “[It] felt more like a conversation in some ways with the audience (even though I only asked one question)—also a bit more like a performance” (pr1). While all presenters could prepare as much as they liked during the rehearsal phase, each indicated that they improvised during their presentations more than anticipated (AVG=4.67, SD=0.33). Presenters reported that this allowed them to follow live feedback and the general mood of the audience to direct their content or even alter their mapping. “I also felt more freedom to change what I was doing as I went along instead of just reading prepared materials” (pr3).

**A Living Document.** One pivotal difference between traditional presentations and Clio is the nature of the final artifact. This was noted by presenters, audience and commentators. “You’re not making slides, which is usually what I focus on when preparing a talk. Instead you’re focusing on the presentation and performance. It’s
not static and can evolve over time. It’s a living document” (pr2). Many participants (pr1–3, pr5, a1–a6, a9, c1, c3) noted Clio made a similar difference in the final product. “Now that I’ve seen this, slides seem more like notes that you can give someone. This is different... it’s a performance and a presentation. It’s alive” (c3). Clio does not produce a final artifact such as slides that could be given to people for review, but a recording of the talk could be distributed. The semi-ephemeral nature of the final artifact produced by Clio parallels the ephemeral nature of semi-extemporaneous presentations given that both are meant to be experienced in the present moment.

Principal Modality In Section 6.6 we noted that similar operations were grouped together into the “arrange” operation, and that the interactions used in the mapping process change to match the interactions used to operate the authoring menu. Similarly, we consistently mapped the interaction mapped to the “text display” operation to manipulate text-based manipulations of the authoring menu (e.g. inputting keywords for voice operations). None of the participants commented or even seemed to notice this, suggesting that this adaptation created a transparent, natural authoring experience. Design of future multimodal systems should consider this collection of operations as a “principal modality”, meaning that the interaction mapped to these operation is performed by users to control the authoring interface itself. Consistency in mapping of the interaction associated with “arrange” and “text display” with the interaction used to operate appropriate elements of the authoring menu maintains immersion in the authoring environment. Future systems incorporating mixed modes of input should consider consistently mapping these operations.
6.11 Limitations and Future Work

**Limitations and Future Work**

**Scalability of Presentation Style.** While we did not explicitly give a time limit for our presentations, participants still opted to give shorter presentations. We collected speculative audience feedback regarding preferences for presentation style if presentations were longer, but knowing exactly how Clio would be used if presentations were scaled to long format is outside the scope of this initial investigation. Understanding how the RtF framework and direct manipulation immersive authoring would affect communication in longer format presentation is an ample avenue for further work.

**Novelty Effects.** While participants in our studies were generally enthusiastic about the potential for incorporating direct manipulation of media into VC communication, the benefits and preferences expressed during the study may change over long-term use of the system. It’s difficult to disentangle how much enthusiasm expressed by participants is accounted by system benefits, and how much is a product of novelty effects. While the findings expressed in this work serve as a foundation for guiding the design of future direct-manipulation VC systems, studying the long-term effects of incorporating such a system into long-term practice remains the subject of future work. One way to approach questions of novelty effects is to deploy a longitudinal study, comparing how behavior and attitudes toward direction manipulation in VC systems shifts with extended use. Conducting such a study would be an ample subject for further research into direct-manipulation VC systems.

**Direct Comparison Study.** Since all participants in our validation studies had significant prior experience with slide-based VC presentations, we did not perform a direct comparison between Clio and traditional slide VC presentations. Our study instead relied on participants prior experiences with traditional slide-based VC pre-
sentations to compare with their experience using Clio. While this approach was sufficient for the purposes of this work and mitigated the demands placed upon our participants, a direct comparison study could yield additional insights into the differences between these two presentation styles. Conducting such a study is a topic for future investigation into VC system design.

**Expanded Operation Vocabulary.** As mentioned in Section 6, we only supported the most commonly cited operations elicited during our formative study. Subsequently, we were not able to revisit all of our participants from the formative study in our validation study. Many of these suggested operations are so rich that they would require their own study to completely explore. Proposed operations included 3D manipulation using voice and gesture and augmenting physical objects with virtual media (e.g. superimposing a virtual photograph over a blank index card). One approach could be to expand the current workflow proposed in this paper. However, tangible and shape changing interfaces present another potential approach to supporting these embodied interactions which encourages creativity and playfulness [46]. Future work will explore these avenues and how they interface with our proposed rehearsal-to-performance workflow.

**Applications Beyond Presentations.** Participants suggested a variety of use cases beyond semi-extemporaneous virtual presentations for employing a system similar to Clio. “*It* felt like it would be super useful when navigating nonlinear media, like a Miro board, or while reviewing something where people will have questions and might need to jump back to something that was discussed earlier” (pr1). Other suggested applications included drawing lessons, remote art studio sessions, virtual classrooms for children, remote litigation proceedings (depositions, trials, etc.), client pitches, storyboard and design brainstorming, musicians and composer collaboration or performance, ASL interpretation, and chatting on social media. Of the future ap-
Limitations and Future Work

Multimodal Telepresence

Applications suggested by our presenters, audience, and commentators, education (87%) and creative applications (82%) were by far the most commonly suggested.

Artifact Production. It was noted by our participants that systems such as Clio produce a 'living document' which focuses the performer's attention on the presentation itself, instead of the artifact of the production (e.g. slides and written material). While this may be preferable in some contexts, certain use-cases necessitate the production of detailed content, visuals, written materials, or other artifacts to help audience members better understand the topic. For example, students may want a copy of slide material from a class lecture to use in studying. One approach to alleviating this concern is to incorporate artifact production into direct manipulation VC systems. For example, systems such as Clio could produce video content, or transcripts of the presentation to be reviewed by audience members at a later time. Other artifacts such as drawings or notes displayed as on-screen text during a performance could be made available after a performance has concluded. Some of these features are included as accessibility tools in commercial VC systems [150]. Transcripts and post-session recordings are available in Teams, for example, as accessibility feature to help improve access to VC sessions to people who are d/Deaf or hard of hearing. Future work in this domain could explore how artifacts can be generated during performances to increase the accessibility and interoperability of session topics.

Accessibility. This paper explored the potential benefits of direct-manipulation in VC systems for able-bodied people, but many open questions remain regarding how a system like Clio would scale to users with different abilities. Gesture control, for example, could be difficult to perform for a person with motor difficulties, but speech control may be easier, and enable more expressive control of VC systems than mouse and keyboard. Similarly, understanding and placing visual media onscreen
presents unique challenges for a person who is blind or visually impaired, but voice commands may be augmented to enable accessible direct manipulation, or wearable devices such as smartwatches could be employed to provide haptic feedback. Some prior work in this domain is evident in the literature, but a dedicated study and system design is absent [52] [50] [134] [150] How people with different abilities and preferences navigate a direct manipulation system such as Clio is an ample and complicated topic demanding a unique investigation to fully understand. We intend to perform a follow-up investigation to this work exploring how direct manipulation of VC media may be enabled by incorporating multi-modality, and how these tools affect the accessibility of VC communication.

Section 6.12

Conclusion

Reimagining remote presentations to take advantage of immersive and directly manipulable environments requires us to rethink not only how we give presentations, but what a presentation should be. Adapting traditional tools like slides for virtual presentations results in a mental model where the artifact of the talk (slides) is the primary focus of preparation, instead of communicating ideas through speech. As demonstrated in our user study, Clio supports creating a ‘living document’ that centers communication and connection with the audience, as opposed to producing an artifact. We also observed that the playfulness of Clio helped to support extemporization because playfulness leaves room for error without sacrificing audience engagement. This playfulness is partially made possible by leveraging the multi-modal capabilities of virtual environments such as machine-learning approaches to gesture and body-pose tracking, as well as voice commands. We also observed that direct manipulation created an immersion so seamless, it enabled presenters to inter-
act with static images using voice and gestures as if they were animated characters. Our rehearsal-to-performance workflow suggests a new approach to presentation authoring and style of presentation deliverance that we believe will inspire future VC environment development.
Accessible Virtual Spaces Using Mixed-Reality Techniques

As video conferencing (VC) has become necessary for many professional, educational, and social tasks, people who are d/Deaf and hard of hearing (d/DHH) face distinct accessibility barriers. We conducted studies to understand the challenges faced by d/DHH people during VCs and found that they struggled to easily present or communicate effectively due to accessibility limitations of VC platforms. These limitations include the lack of tools for d/DHH speakers to discreetly communicate their accommodation needs to the group. Based on these findings, we prototyped a suite of tools, called Erato that enables d/DHH speakers to be aware of their performance while speaking and remind participants of proper etiquette. We evaluated Erato by running a mock classroom case study over VC for three sessions. All participants felt more confident in their speaking ability and paid closer attention to making the classroom more inclusive while using our tool. We share implications of these results for the design of VC interfaces and human-the-loop assistive systems that can support users who are d/DHH to communicate effectively and advocate for their accessibility needs.
Recent studies indicate that hearing ability correlates with employment success for the 20% of adults in the USA who are Deaf or Hard of Hearing (d/DHH) [9]. We use the term d/DHH to refer to three groups; “deaf” with a lowercase “d” to reflect the physical condition of deafness, “Deaf” with a capital D to refer to people who identify as culturally Deaf, and “Hard of Hearing” referring to a wide range of conditions resulting in mild to moderate hearing loss. People who are d/DHH are employed at a rate 38% less than their peers who are non-d/DHH with comparable experience and education. Furthermore, employees who are d/DHH receive 34% lower salaries and wages on average compared to their non-d/DHH coworkers [40]. Previous investigations have illuminated a crucial element behind these inequities: effective communication in small-groups is vital to the success of employees with d/DHH [1, 31] and employees with d/DHH often struggle when communicating with their peers who are not d/DHH in professional environments [23]. Communication difficulties in small-groups for people with d/DHH extend beyond the workplace, including classrooms [44], brain-storming [12], and social engagements [32].

In recent years, many of these small-group professionals and educational activities have moved online to remote video conferencing (VC) platforms such as Teams, Meet, and Zoom [22]. Prior research has investigated how to communicate to participants who are d/DHH on VC platforms such as improved captioning and support for signers [26, 34]. Little work, however, has investigated how systems can be designed to support speakers or presenters who are d/DHH and verbal. Recent commercial and research tools [1] have explored the use of automation to provide real-time support for VC presenters or feedback for presenters during rehearsal [9, 21]. However, the
extent to which these tools are effective at supporting presenters who are d/DHH during real-time VC presentations or during other activities which require small-group communication with mixed-ability groups is an open question.

In this paper, we explore the design of enhancements for VC platforms to better support people who are d/DHH while presenting or speaking in small group discussions. We first conducted a formative fly-on-the-wall study to examine the accessibility limitations of current VC platforms to support presenters who are d/DHH. We found d/DHH participants prefer to conduct a variety of tasks on VC platforms versus in-person spaces because they allow for the use of captions, recordings, and other tools that help them understand the discussion material. However, we also discovered that supporting participants in VC spaces who are d/DHH and wish to speak still face various difficulties including discreetly communicating and enforcing their accommodation needs, regulating their own speech speed and volume, as well as regulating the speech speed and volume of other presenters. In addition, we found that VC meeting hosts often are unaware of the accommodation needs of participants who are d/DHH during small-group VC discussions.

To address these limitations, we prototyped Erato, a suite of tools to support inclusive video-based communication collected into a Chrome Extension. Erato employs a data-driven approach to provide real-time feedback for speakers who are d/DHH during VC sessions, as well as a socio-technical support system to allow participants who are d/DHH in VC discussions to discreetly express and remind participants of their accommodation needs. Our system also provides a subset of tools for meeting hosts to provide at-a-glance updates on the participation of VC discussion members, as well as methods to intervene and make the VC session more inclusive. We deployed Erato in a three-session longitudinal case study examining both d/DHH and non-d/DHH participant behaviors over time and validity across common tasks conducted on VC.
platforms. The principle contributions of this work include:

- A series of design considerations for supporting d/DHH speakers in VC environments informed by synthesizing the results of prior work as well as our formative surveys, interviews, and formative studies of current VC platform’s accessibility limitations.

- The design and implementation of a suit of tools, called Erato, providing data-driven socio-technical support for presenters who are d/DHH.

- A longitudinal case study evaluation of Erato which showed that inclusive VC system etiquette could be quickly learned by participants and maintained across sessions, improving the accessibility of these systems.

We found that anonymously communicating accommodation needs were embraced by d/DHH and non-d/DHH participants alike, and that accommodation requests for etiquette behaviors were easily incorporated into the communication habits of all participants. Furthermore, these behaviors carried over across sessions with little enforcement or reminder required on behalf of the participants. In addition, speaking tools were not only used by speakers with d/DHH to visualize the rate and volume of their speech, but these same tools could also be used to communicate anonymously to other participants how best to speak for optimal caption performance and lip-reading efficacy. We also discovered how these tools could be applied to not only mitigate inequities of other disabilities, but also the discrimination experienced in VCs by other minorities including women, queer people, and people of color.
Our research intersects with many existing bodies of interdisciplinary literature including assistive technologies (7.2.1), video conferencing accessibility tools (7.2.2), and explorations of accessibility support for presenters in video conferences (7.2.3).

7.2.1. Audio-Based Assistive Communication Technologies

Significant prior research has investigated how technology can be designed to support people with disabilities in a variety of tasks, as well as how technology itself can be designed to be more inclusive and accessible [20]. A common approach of these assistive technologies is to substitute one sense for another, such as replicating the experience of sound using haptic feedback. For example, as we discussed earlier in Chapter 3 of this thesis, TangibleCircuits used haptic and audio feedback to guide people with disabilities in prototyping hardware computing circuits [52]. Ubiquitous computing and wearable technologies have also been developed to broaden types and combinations of feedback that can be supported in assistive devices [82] [46]. Other approaches explore using smartphone displays to communicate non-speech sounds because these ambient sounds provide safety, social, and critical environmental information [154] [164]. Similarly, Goodman et al. use smartwatches to evaluate ambient environmental sounds and provide visual and tactile feedback to users who are d/DHH [81].

Work investigating sound quality and how it can be effectively represented as captions or communicated using multimodal approaches is a vibrant, active field of continuing research. Differentiation of individual sounds from the surrounding environment to amplify the volume of crucial sounds has been explored [39] as well as communicating sound isolation using captions or tactile approaches [135] [27] [95].
Dynamic captions that adapt to the user’s field of view have been developed by analyzing eye movement and gazing patterns [27]. Similarly, gaze-adaptive captions are re-positioned based on the viewer’s gaze and the objects present on the screen [120]. Information-rich subtitles have also been explored which color code voice volume and speech information to help illustrate the mood/tone of the speaker [95]. Automated speech recognition (ASR) captions used ASR technologies to generate captions for online videos in near-real time [135] [215]. Besides providing captions for English, captions for American Sign Language (ASL) have also been proposed [198]. Other projects propose improving the reliability of captions by communicating the ASR systems confidence in its own captioning [18]. Often in VC sessions, users who are d/DHH will be accompanied by an ASL interpreter to help facilitate communication. Seita explored placement of the interpreters’ window at the bottom of the screen in VC sessions, similar to where traditional captions would be displayed [207]. Other work has explored the social implications of incorporating interpreters into social software systems [140]. Automating the role of an ASL interpreter is also an active area of research. Elliot et al. explore the affordances of commodity smartphone technologies with their prototype app to accomplish smooth communication between a d/DHH ASL user and a non-d/DHH user unfamiliar with ASL [64]. Mixed reality (MR) and virtual reality (VR) methods have been proposed which insert visual cues (highlights on the speaker’s table or augmenting the field of view with visual queues) to indicate the current speaker [77]. Head-mounted display (HMD) technologies have also been developed to project an augmented reality (AR) sign language interpreter just outside the view of the user while they are watching a television program to provide automated ASL communication [243]. AR glasses or smart glasses have also been explored to produce real-time captions for users who are d/DHH [103] [181]. Other approaches employ novel display interfaces to bridge accessibility communic-
tion gaps [47]. Kushalnagar et al. employed projectors to directly display real-time captions on top of the current speaker’s head in a classroom environment [30]. Other approaches investigate how to use smart-glasses or projectors that insert captions beside the presenter [190] [122], or provide support for people who are not-d/DHH to communicate using ASL [85].

7.2.2. Video Conferencing Accessibility Tools

Video Conferencing has recently seen a broad expansion in everyday use-cases ranging from classroom, seminars, panels, corporate meetings, and casual social interactions [50]. This is especially prescient following the COVID-19 outbreak of early 2020, where numerous daily activities hastily moved to VC platforms [121] [63] [153]. As a result, systems expanding the capabilities of video conferencing have emerged. CLIO, discussed in Chapter 6 of this thesis, incorporates gesture detection and voice commands to enable a presenter to manipulate on-screen media and text [48]. In addition, many accessibility limitations of VC platforms have been documented, such as the importance of users who are d/DHH to see the full body and face of the presenters [121], to have fewer visual distractions during the session to mitigate cognitive load [153], the extent and limitations of incorporating automated alt-text generation [134], and to have access to reliable and rich captions [63] or interpreters when using ASL [121]. Many techniques have been proposed to address these accessibility concerns, including tools which use machine vision to optimize the furniture arrangement and body placements of users so their bodies and faces can be clearly seen [121]. To reduce the heavy cognitive workload of users who are d/DHH, researchers have explored consolidating points of interest in VCs to minimize visual distraction [167] [11] [34] [25]. Virtual classrooms and e-learning systems have also been proposed for their ability to bridge accessibility needs by leveraging the interaction capabilities of VR [41] [55]. Other approaches propose visually augmenting VCs, to reduce cognitive
overload problem that originates from shifting attention between stimuli [168]. Other techniques explore how users with d/DHH can effectively insert their ASL interpreter into the meeting [199] or allow human-in-the-loop approaches to improve ASR and captions [17] [18].

VC systems have a unique set of challenges to overcome regarding cognitive overload due to the breadth of visual stimuli common to VC platforms [185]. Also, new difficulties are introduced such as people with d/DHH being unable to read lips when participants have their cameras turned off or the system automatically hides the person speaking [228]). Significant prior literature documents these concerns and proposes guidelines for etiquette which could make VC activities more inclusive for people who are d/DHH [185] [198] [23] [123] [4] [244] [228] [208] [153]. Kushalagar et al. suggest that the participants help themselves by designating a person to monitor the chatbox and read it when someone sends something or turns on video only under specific conditions [123]. After conducting a meta-analysis of current literature, Bouzid et al. suggest improving the accessibility by focusing on orienting the camera towards the face, normalizing the incorporation of ASL translators, and redesigning the visual interface [23]. This work also highlights common pain-points engaging accessibility features, such as poor internet bandwidth or the cognitive overload of text-sparse media. Vogler et al. on the other hand, suggest improving accessibility by standardizing software and hardware protocols that account for specific needs [244].

7.2.3. Accessibility Support for Video Conferencing Presenters

While prior work has explored making video conferencing more accessible for people with d/DHH, it focuses exclusively on making VC platforms more usable by people with d/DHH as audience members, and not as speakers, presenters, or session hosts. Prior work explores dynamics in mixed-ability VC environments [150]. Other preliminary work focuses on those fluent with ASL or with access to an interpreter [199]
Most closely related to our interests are works which consider supporting presenters, hosts, and speakers who are d/DHH who might not have access to interpreters, nor be fluent in ASL. The work of Rusnák et al. suggests adding features such as “raise hands/give floor” functions in their VC prototype so that the requests from the audience could be spotted easily by speakers who are d/DHH. Similarly, CollabAll allows participants to suggest classroom etiquette violations and politely interrupt the instructor for any needs. Seita et al. remotely co-designed accessible features for VCs and found that notifications system influences speaking behaviour and suggest providing options for communication modalities and prioritizing d/DHH participants’ communication preferences. Prior research also indicates that automated detection tools and notification mechanisms significantly regulate presenters’ behaviour. McDonnell et al. suggest that information-rich captions enabled by adding automatic tools checking speech rate and volume positively impacted the communication of both d/DHH and non-d/DHH users during VC sessions.

These prior works suggest that VC platforms could provide significant support for people who are d/DHH speaking and presenting information during a session. In this work, we survey current tools and commercial products for accessibility features, and collect relevant features into a single source, we consolidate visual stimuli and provide data to session hosts to allow them to make decisions regarding the accessibility of their session aiming for a decrease in cognitive load. In addition, we co-design new tools to provide means to remind session participants of inclusive etiquette and enable session participants to produce their own accessibility accommodation guidelines according to the needs of the group members. Finally, with these tools consolidated into a single suite we are able to produce rich qualitative insights into their effectiveness and usability by people who are d/DHH in a usability study.
To understand how current VC platforms support the needs of participants who are d/DHH, we designed a study consisting of surveys, semi-structured interviews, and a fly-on-the-wall observational use-case study. While prior work provided motivation for the problem space, we wanted to better understand of the nuances of the actions and behaviors around pre-existing tools and elucidate actionable problems and decided to use an observational approach [225]. Our investigation probes general pain-points and benefits of VC platforms for d/DHH users, compares accessibility features available in current commercial platforms, how these platforms mediate and affect communication for mixed-ability groups, and how experiences using VCs change between contexts for d/DHH people. We chose to include non-d/DHH participants and mixed-ability populations in our formative study to mimic the mixed abilities conditions common to real world VC use.

7.3.1. Participants

Of the 26 responses to our survey, 14 self-identified as having one or more disabilities, 7 of which self-identified as d/DHH. Respondents ranged in age from 18 - 44 with a median age of 26. We interviewed all 14 respondents who self-reported having a disability and experience with at least one VC platform. Of these, 7 identified as cisgender men, 5 identified as cisgender women, 1 identified as two-spirit, 1 identified as demi-male. The 4 of the 7 interviewees who disclosed they are d/DHH also reported having additional disabilities; ADHD, Autism Spectrum Disorders (ASD), and visual difficulties. Participants occupations were mostly students (71%) with 14% reporting they worked in education, 7% worked in other fields, and 7% reported being unemployed. Participants for our mock-classroom study were chosen based upon their
experience in remote classrooms conducted via VCs. Our instructor was an experienced educator who was invited based upon their expertise in teaching as well as their experience teaching both physical and remote classrooms via VCs with mixed-ability student populations. The inclusion criteria for the interview was based on participants’ previous experience using VCs with disabilities for a variety of use-cases. All d/DHH participants with disabilities were interviewed and d/DHH participants who had previous experience in both VC and physical meeting spaces were prioritized for the live study. Non-d/DHH participants were also included in our survey/interviews because we wanted to understand more broadly how VC systems could be made more inclusive and accessible, and many of the respondents self-reported disabilities other than d/DHH. In addition, we chose non-d/DHH participants for inclusion in our formative study who had prior experience with both VC and physical meeting environments, and we selected an event mix of participants who had prior experience in mixed-ability spaces and those who did not. A table detailing characteristics and demographics of participants included in the live study can be seen in Table 7.1 and a full table of characteristics for all participants included in the survey and interview study can be found in Appendix E.

### 7.3.2. Methodologies and Procedure

Prior to participation in any portion of the study, participants were provided informed consent to participate as stipulated by our institutions IRB, including participants’ rights to refuse participation in any portion of the study that made them uncomfortable, or quit participation in the study at any time. In addition, participants were informed that their identity would be provided anonymity in any subsequent publication resulting from the study, and any quotes would be attributed to a moniker such as P1, P2, etc. The process of informed consent was completed at each portion of the study, with an option to leave the study and remove their data from the col-
Table 7.1: Participant table for members selected from survey and interview participants for the mock-classroom formative study. Participant noted with a * symbol indicates that this person served as the session host for the study. Fields noted as N/A indicate participant declined to answer this survey question. Full table of participant characteristics included in Survey and Interview portions of the study can be found in Appendix E.

<table>
<thead>
<tr>
<th></th>
<th>Self-Described Disability</th>
<th>Age</th>
<th>Gender Identity</th>
<th>Occupation</th>
</tr>
</thead>
<tbody>
<tr>
<td>FP1</td>
<td>Progressive hearing loss</td>
<td>20</td>
<td>Cis-Man</td>
<td>Student</td>
</tr>
<tr>
<td>FP2</td>
<td>ADHD</td>
<td>21</td>
<td>Demi-Male</td>
<td>Student</td>
</tr>
<tr>
<td>FP2</td>
<td>ASD</td>
<td>26</td>
<td>Cis-Woman</td>
<td>Student</td>
</tr>
<tr>
<td>FP4</td>
<td>None</td>
<td>N/A</td>
<td>Cis-Woman</td>
<td>University Instructor</td>
</tr>
<tr>
<td>FP5</td>
<td>bilateral hard of hearing since birth, low vision</td>
<td>36</td>
<td>Cis-Man</td>
<td>Student</td>
</tr>
</tbody>
</table>

Selection at any time. We distributed a survey online through social media as well as in partnership with the Accessibility Accommodation Office at our institution. Respondents were asked 7 demographic questions including optional disclosure of any self-described disabilities, 12 short answer questions on their previous experience using VCs including contexts, use cases, and general impressions, followed by 15 Likert scale questions. From the collected responses to our survey we selected 14 participants (7 who self-identified as d/DHH) for follow-up semi-structured interviews to further expand on responses to the survey. Each interview lasted approximately 60 minutes. All interviews were recorded and transcribed for research purposes except for 4 participants who declined to be recorded. In these cases, hand-written notes were taken of the conversation.

We selected the use-case of a remote classroom experience as it was suggested by the majority of our participants (88%) as a real-world context in the survey feedback. The strong prevalence of this suggested use-case likely occurred as 85% of the survey participants indicated that they were students or working in education. We recruited
from both online forums as well as working in coordination with our institutions Office of Accessibility, which would account for the strong student and education demographic presence. Given the deep participant familiarity with the remote classroom context, it was suitable as an environment for our observational study. 5 participants (1 non-d/DHH instructor, 2 d/DHH students, and 2 non-d/DHH students) were selected from our interviewees. Each participant was assigned a short reading (less than 2 pages) from a text on psychology, and asked to present a 3-5 minute summary of the paper contents followed by 3 questions to lead a group discussion on the presentation content. On the day of the study, our instructor facilitated each participant presenting their summary, followed by the presenter leading a discussion based on their presentation with the whole group. After each speaker we distributed a questionnaire asking each participant questions regarding their social experience during the presentation and discussion. Researchers turned their camera and microphones off for the duration of the presentation, observing, taking notes, and posting links to appropriate questionnaires following each presentation. After the final presentation, participants answered a final questionnaire and completed an exit interview with our research team. During exit interviews, researchers discussed questionnaire feedback and observations with the participant, clarifying and elaborating as needed. Exit interviews were scheduled individually based upon participant availability. Interviews were conducted by 3 members of the research team who also performed analysis of the qualitative data together following the conclusion of the interview process. The live VC study took 60 minutes to conduct and the exit interviews each took between 15 and 60 minutes.

7.3.3. Analysis

Both the survey responses and interview transcriptions were coded for common themes and recurring pain-points mentioned by participants using a grounded methodology.
Interviews were conducted by 3 researchers who, after concluding all interviews, gathered to perform a cursory analysis of the data and collect a base set of codes from the qualitative results. From these base codes, intermediate codes were agglomerated from these initial codings by consensus from the 3 researchers. For example, intermediate codes included social awkwardness, feature limitations, feature preferences, and socio-technical dynamics. We then recruited interested respondents for a fly-on-the-wall use-case study to better understand the social dynamics and mental model arising from mixed-ability VC systems. We also collected accessibility support feature preferences and pain points for 3 common VC platforms through our surveys and interviews, extracting common patterns of tension reported by our participants and looked for evidence of these limitations in VC systems. We make special note of whether or not each tool is hidden behind a paywall given the financial inaccessibility findings reported by respondents. Complete findings can be seen in Appendix F.

7.3.4. Findings

*Physical vs VC Spaces.* Participants expressed that they preferred VC platforms to physical spaces for lecture-based classroom activities for several reasons. First, VC platforms allow the use of captions which reduce reliance on lip reading for communication. In addition, since VC platforms are video-based, they could be recorded and watched again later in case the student missed part of the lecture due to volume or lip-reading difficulties. The ability to adjust volume was also valued by participants since it provided an aural control affordance not possible in physical spaces. For students who rely on lip-reading, VC platforms encourage participants to face the camera and due to the form factor of most laptops, cameras used in VC communication are generally positioned towards the face. The positioning of the camera increases the likelihood that participants who are d/DHH will be able to clearly see the face and lips of the speaker, thus reducing the difficulty of lip reading that might result from
7.3 Formative Study

physical spaces. For these reasons, all participants who are d/DHH prefer the use of VC platforms over physical spaces for lecture-based classroom activities.

**Pain Points.** A key pain point of current VC systems is the lack of a robust system for instantiating accessibility accommodation requests during the session. VC platforms lack a method for people who are d/DHH to request and have their accessibility needs met by a host or other session members. Participants reported that they have to publicly request that their accessibility needs be met, which is embarrassing and uncomfortable, often resulting in participants fore-going any accommodation at all. Our expert instructor corroborated this result, detailing how they would try to accommodate as many requests as possible, especially if a student contacted them in advance, but often these requests would get “buried in email and I wouldn’t be able to read them in time” (FP4). Furthermore, since remote VC platforms were so rapidly adopted during 2020, common practices for making discussion sessions accessible are still not widely known or enforced as expected social etiquette.

While VC platforms were generally preferred for lecture-based classroom activities, several pain points still arose when discussion-based, brainstorming or activities requiring multiple participants to speak took place via remote VC. Participants reported that monitoring the “text chat and captions simultaneously was exhausting” (FP5), and resulted in students with d/DHH often having to “re-watch the lecture several times to account for the parts of discussion [they] missed while reviewing chat in the text box” (FP1). Volume and speech rate also varied from person to person in VC based discussions, and controlling the volume for each individual user is significantly taxing. When speaking themselves, participants with d/DHH expressed anxiety around their own volume level or speech speed while speaking, indicating a mode of communication regarding oral speech parameters was missing from current VC tool suits. Insecurities around speaking were a common theme among all
participants with d/DHH, expressing that the usual anxieties of public speaking are compounded by difficulties hearing yourself speak.

The instructor also expressed a need for several tools which would help them better facilitate VC sessions. Data regarding who had spoken most recently, how much session participants had spoken, or who hadn’t spoken at all, were indicated as data that would be valuable to the instructor in ensuring their VC classroom was inclusive and accessible. Reminders to record or turn on captions at the beginning of the session was also expressed as a valuable feature for the instructor, since these are tools which are available but often go unused. Finally, despite the presence of many accessibility tools such as captioning and video recording/transcription already being commonly included in commercial VC platforms, many of these tools are hidden behind a pay-wall. This creates a barrier of financial inaccessibility, and even when these tools are not hidden behind a pay-wall, many educators and meeting hosts are unaware that accessibility features exist, or the crucial role these tools play in making VC platforms usable by all people.

7.3.5. Design Goals

Our formative study results along with insights from prior work indicate that the limitations in accessibility evident in VC platforms extends beyond applicability to the classroom use-case explored in our study. Developing additional tools to complement built-in VC capabilities is necessary to support speakers who are d/DHH, provide necessary data to VC meeting hosts to ensure that their sessions are accessible and inclusive, as well as alter the etiquette of VC sessions to be inclusive in a variety of contexts. Based on the above findings from our formative studies, we synthesize a series of design goals to guide the development of our tool suite and subsequent research.

- **G1.** Facilitate communication of accessibility accommodation needs, and sup-
port reminders of these needs sufficient to alter behavior and instate common accessible and inclusive etiquette.

- **G2.** Support speakers in VC sessions who are d/DHH by providing real-time data-driven feedback.

- **G3.** Provide meeting hosts with data-driven tools to inform their instruction of the class and ensure the space remains inclusive.

- **G4.** Evaluate the accessibility and usability of speech support tools and explore how these tools can be complemented or improved to better meet the needs of end-users who are d/DHH.

---

**Erato System Design**

To address the above considerations, we designed and implemented Erato, a chrome extension for improving inclusiveness in VCs by connecting participants and hosts for accessibility needs of d/DHH participants and enabling self-improvement with automated tools. Erato is a Chrome Extension that can be installed on a Chrome Browser with version 88 and up. There are two roles when using the software: a participant role and a session host role, each with a set of fundamental features intended to expand the inclusiveness of the VC session. Features incorporated into Erato were agglomerated from multiple sources to create a complete suite of tools for evaluating effectiveness of their support for speakers who are d/DHH (See G4 in Section 7.3.4). These sources include tools present in commercial software and state-of-the-art research (which we recreated using prior literature) as well as open source tools. The remaining tools are novel and prototyped according to feedback received during our formative study.
7.4 Erato System Design

We group Erato’s features into three groups following our design goals mentioned in Section 7.3.5 (where appropriate, we include the source of each tool in the description below): accessibility accommodation requests and etiquette guide in Figure 7.1 (G1), speaker support features in Figure 7.2 (G2), and host exclusive features in Figure 7.3 (G3).

7.4.1. Accessibility Accommodation Requests/Etiquette Guide

Erato’s Accessibility Accommodation Requests/Etiquette Guide establishes a channel of communication for meeting hosts and participants intended to communally practice and re-enforce accessibility accommodation requests. The etiquette guide functions on the principle of nudges and persistent reminders by bring the accommodation requests to everyone’s attention [16]. Using the anonymous accommodation request interface, participants and the host are able to submit personalized accessibility accommodation requests at any time during the meeting. The requests that participants and hosts enter are anonymous to mitigate any potential hesitancy or embarrassment that may result from making accommodation requests. To ensure the submitted requests are respectful to every participant, practical for the needs of the meeting, and do not contain harmful language, they are reviewed by the host before being presented to session participants. The host can approve, delete, or archive incoming accommodation requests using the host accommodation interface. Approving a request by the session hosts adds the accommodation request to the etiquette guide, visible to all participants. Any requests deleted from the request log by the host are removed from the system. This is intended to prevent trolling or abuse of the system. Along with this presentation, all participants and hosts will receive a one-time notification at the top of the extension to bring this request to the group’s attention.

To maintain adherence to the approved accessibility accommodations comprising
7.4 Erato System Design

Accessible Telepresence

Figure 7.1: Overview of Erato Accessibility Accommodation Tools: (A) Participants and hosts submit anonymous requests in the text input. (B) Submitted requests are listed for the host to approve or reject. (C) Buttons allowing hosts to approve or reject submitted requests. (D) Approved requests are listed for all participants. (E) Participants and hosts can report a violation of the request, the number next to the bell indicates the count of violations reported by participants during the session. (F) Recently approved requests and reported etiquette violations will appear as separate notifications in the Notification Center.
the etiquette guide, Erato supports a communal reinforcement notification system. Whenever a participant or host want to bring attention or remind the session of a particular request, they are encouraged to use the “bell” (report) button to remind session participants of the importance of following the accessibility accommodation request. The Erato system stores the number of incident reports (i.e. report button clicks) for each accessibility accommodation, and adjusts system behavior to subtly emphasize adherence to accessibility requests with a higher rate of incident report. Once a report button has been clicked, a slide-in notification reminds all session participants of the corresponding accommodation. If multiple accommodation requites are present in the etiquette guide, Erato orders them from most reported to least. The guide is also color coded according to which etiquette items have been reported most recently. Items reported within the past 5 minutes are colored red, with opacity of the red coloring increasing with each subsequent report triggered within a 5 minute interval. This red background opacity decreases for each minute after a 5 minute interval without subsequent report button clicks, returning to white after after 15 have elapsed without report. The report feature is also anonymous as the person reporting the violation is not named in notifications, nor is their request stored by the system in any way. It is also worth noting that this notification is not sent to a specific participant, which not only underlines the requests to all participants but also reduces the pressure on any member of the session not directly adhering to the accommodation request.

All approved accommodation requests are stored in the Accessible Accommodation sections until the database is flushed. This forms a small etiquette guide for a subsequent series of meetings and relieves repetitive submission of the request whenever the meeting begins. The requests are stored in order of the report count with the consideration of the time: the request will be ranked higher if it has a large number
7.4 Erato System Design

7.4.2. Speaker Support Features

In addition to the etiquette guide and behavior reinforcement system, session participants are provided a collection of tools to encourage the inclusiveness of meeting
sessions and to assist with their speaking. Session participants can check their real-time spoken volume, monitor their speech speed throughout the session, and listen to their voice using direct audio playback. The direct playback feature was specified as a common need by P3 of our formative study because VC systems often make it difficult for people using hearing aids or cochlear implants to hear themselves speak. The playback feature allows participants to play their own spoken audio through the aux cable directly connected to their assistive devices.

Participants and hosts can monitor their real-time volume by glancing at the volume meter that has three magnitudes: the green bar in the middle, which indicates the volume is comfortable for others to hear and the two yellow bars on each side, which shows the volume is too soft or too loud. Similarly, a speech speed meter is located adjacent to the volume meter and provides similar feedback to the volume meter. The speech volume and speed recognition systems use Mozilla speech recognition API to measure the number of words spoken in a sentence divided by the total time spoken, producing an approximate calculation of the average words spoken per minute. We filter out sentences that have few words or take a short time to speak to eliminate detection of filler words and sounds. A separate notification is sent by the system if a substantial number of filler words are detected during a given session. Default values were informed by prior literature on ideal speech speed and volume, as well as ranges employed by automated speech volume and speed feedback tools incorporated into commercial products such as Microsoft Cameo for PowerPoint [209]. Speech volume and speed magnitude ranges are adjusted when audience members use the speech speed and human volume buttons located below each respective tool. Pressing the increase or decrease volume or speed buttons shifts the magnitude ranges in the respective direction, allowing audience members to directly communicate impedance desires to the speaker. This design is intended to provide a
human-in-the-loop adjustment to the automated feedback tools, allowing speakers to quickly and peripherally receive feedback on their speaking from the audience.

In addition to glancing at the meter, speakers can receive at-a-glance textual feedback on the top of this section with the same colour code as the volume meter. Green indicates that they are speaking within the desired volume and speed ranges and yellow indicates that the speaker may need to adjust their volume or speed. Participants and hosts can get more specific and accurate feedback by looking at the numerical value displayed on the volume and speech meters. This numerical information is brief to avoid distracting the presenter, and is intended to provide complete data transparency that could be useful for speech preparation. The color coding and magnitude design choices are intended to reduce the cognitive load on the speaker, and provide needed information at-a-glance.

The **Playback** feature plays the audio recorded by the microphone synchronously into the participant or host’s earphone. This feature was directly requested during the interviews of our formative study by a d/DHH participant who indicated that using an audio cable directly plugged into their hearing assistive devices gave them greater control over the ability to hear their own voice when speaking. Without such a feature, people using hearing aids often unintentionally taper their volume due to difficulty recognizing their own spoken volume levels. Participants in our formative study noted that they often used third-party software to mitigate this issue, which often was not compatible with different VC systems. We include this feature within Erato to create a seamless collection of tools for supporting speech in VCs.

**Notification Center** gathers all the notifications sent by the participants and hosts into a single centralized location. Notifications are displayed in chronological order of their submission, and repeat message requests (e.g. multiple etiquette reminders generated by clicking the “bell” button, or multiple human volume button
7.4 Erato System Design

clicks) are filtered into a single notification to simplify the display and reduce cognitive load.

7.4.3. Host Exclusive Features

Erato also has features that are exclusive to the meeting host and provide the host with sufficient data to enable their session to be as inclusive as possible. Some tools are toggles which can be requested by session participants, but need to be triggered by the host. For example, triggering the **Keep Text Chat to a Minimum** toggle switch will automatically input into the text box automated replies should participants send text chat in the text messages field. This text message reminds the participant to keep their text chatting to a minimum and orally communicate when appropriate. The automated response will only be visible to the message sender, encouraging the sender. This feature was co-designed during our formative study by participants who were d/DHH and relied on speaker captions during the session to remain informed of session activities. Participants reflected that the cognitive load of reading both the captions and text messages was exhausting, and often resulted in the user becoming lost by splitting their focus between the speech captions, which often disappeared within a few moments, and the text message box. This tool is designed to mitigate cognitive load by giving session participants the opportunity to focus on captioned speech, should they desire.

To complement this, session hosts are provided data-driven tools to inform their understanding of session participant engagement and communication preferences. One panel indicates **Are People Texting or Speaking** to give an intuitive idea of how much the voice or text chat is being employed during the current session. Erato converts the text that participants and hosts typed in the In-Call Message into the number of words input, and compared this to the number of words spoken by participants during the session. This data is reflected in a pie chart to quickly pro-
7.4 Erato System Design

Accessible Telepresence

Figure 7.3: Host Exclusive Features: (A1) Toggle switch to discourage participants from using the text chat to communicate. Once turned on, an automated text notification will appear in the chat display if Erato detects that participants are using the text chat. (B) Participants and hosts are reminded every 40 seconds to turn on their closed captions unless Erato detects that the participant or hast has captions turned on. (C) Hosts can monitor the amount of time each person has spoken during the session, as well as see visual indicators in green of who has spoken most recently. (D) hosts can see the ratio of communication occurring via voice or text usage in this pie chart.
provide the session host at-a-glance data regarding the communication modalities being employed in the session.

Similarly, **How Long Has Everyone Been Speaking**, provides the instructor an intuitive understanding of session participation by individual members of the session. Duration of participant speech is measured and represented using a bar chart for quick comparison of participant activity. The length of the bar chart is standardized by the participant who speaks the most, and visual adjustments are made accordingly. The dashboard also indicates to the host the last three speakers who spoke, indicated by a green highlight. Bar chart order can be ranked in ascending or descending order according to the needs and preferences of the host.

### Section 7.5

**Validation Study**

We conducted a case study to understand the usability and accessibility of speech support tools comprising Erato (See Design Goal G4 in Section 7.3.5). We also wanted to examine how these tools affect social dynamics in 3 common remote VC contexts with mixed-ability groups of people. Our goal was to understand what socio-technical dynamics emerged by incorporating Erato into common VC contexts and how those differed from the behaviors observed during our formative study. We used a longitudinal approach for our study to compare how these dynamics and behaviors evolved over time and across contexts.

**7.5.1. Participants**

We recruited 12 participants for our study: 6 participants (3 d/DHH) participated in all 3 sessions while the other 6 participants (3 d/DHH) were incorporated in pairs for 1 session per pair. This way, each session comprised 8 participants (4 d/DHH), with
6 participating in all 3 sessions and 2 new members participating in only one session. This was designed to mimic real-world conditions since new members joining group activities are common in online VC contexts. All participants were compensated for their time. Participants were recruited by a survey distributed online and through email in partnership with the Accessibility Accommodation Office of our institution. Similar to the questionnaire of our formative study, this initial survey collected demographic information as well as short-answer and Likert data regarding respondent previous experience using VCs. We included 12 participants in a fly-on-the-wall case study based upon criteria pertaining to their experience with VC platforms and mixed-ability groups. Specifically, participants who had extensive experience performing both professional and social tasks on VC platforms were prioritized, especially those who were d/DHH. We intentionally included non-d/DHH participants with a mixture of familiarity and experience working with mixed-ability groups. Many non-d/DHH participants reported being completely unaware of the needs of d/DHH people, and were included to evaluate how inexperienced participants interact with d/DHH participants while using the tool. We chose to include d/DHH participants as well as non-d/DHH participants to mimic real world, mixed-abilities conditions, observe how non-d/DHH participants responded to accommodation requests (especially those unaware of common d/DHH needs and inclusive etiquette), and to evaluate any difficulties non-d/DHH folks may have when using the system. d/DHH participants were prioritized in the selection and scheduling process and non-d/DHH participants meeting the above criteria were included based upon their availability given the schedules of our d/DHH participants. A full description of participant characteristics and demographics can be seen in Table 7.2.
Table 7.2: Participant table for members selected for the Erato prototype case study where P Number indicates anonymized participant number. Participant noted with a * symbol indicates that this person served as the session host for at least one session of the study.

<table>
<thead>
<tr>
<th></th>
<th>Self-Described Disability</th>
<th>Age</th>
<th>Gender Identity</th>
<th>Occupation</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>None</td>
<td>23</td>
<td>Cis-Woman</td>
<td>Software Developer</td>
</tr>
<tr>
<td>P2</td>
<td>Deaf, Monovision</td>
<td>40</td>
<td>Cis-Man</td>
<td>Student</td>
</tr>
<tr>
<td>P3 *</td>
<td>None</td>
<td>32</td>
<td>Cis-Woman</td>
<td>Illustrator and Teacher</td>
</tr>
<tr>
<td>P4</td>
<td>None</td>
<td>28</td>
<td>Cis-Man</td>
<td>Student</td>
</tr>
<tr>
<td>P5</td>
<td>None</td>
<td>32</td>
<td>Cis-Man</td>
<td>Vocalist</td>
</tr>
<tr>
<td>P6 *</td>
<td>Progressive hearing loss</td>
<td>22</td>
<td>Cis-Man</td>
<td>ESL Instructor</td>
</tr>
<tr>
<td>P7 *</td>
<td>Bilateral hard of hearing, since birth</td>
<td>36</td>
<td>Cis-Woman</td>
<td>Unemployed</td>
</tr>
<tr>
<td>P8</td>
<td>Deaf, Low vision/vision problems</td>
<td>57</td>
<td>Trans-Man</td>
<td>Unemployed</td>
</tr>
<tr>
<td>P9</td>
<td>deaf/hard of hearing and chronic illness</td>
<td>24</td>
<td>Cis-Woman</td>
<td>Retail</td>
</tr>
<tr>
<td>P10</td>
<td>ADHD/deaf</td>
<td>19</td>
<td>Trans-Woman</td>
<td>Student</td>
</tr>
<tr>
<td>P11</td>
<td>None</td>
<td>24</td>
<td>Cis-Woman</td>
<td>Software Engineer</td>
</tr>
<tr>
<td>P12</td>
<td>None</td>
<td>30</td>
<td>Gender queer</td>
<td>Writer and English Instructor</td>
</tr>
</tbody>
</table>
7.5 Validation Study

7.5.2. Methodology

We followed the exact same procedure for informed consent as documented in our formative study in adherence to the ethics policy provided by our institution’s IRB (See Section 7.3.2). Each of the study’s 3 sessions took place on a different day over the course of a week to reflect real-world conditions and provide information on how behaviors changed over time. Tasks were chosen based on their common prevalence among the use-cases reported in the qualitative feedback from our formative and case study surveys. These tasks also required participants to use various external tools common in VC contexts such as screen sharing, slide presentation, and whiteboard/collaborative brainstorming tool use. All sessions were recorded with the informed consent of the participants and transcribed for analysis. We averaged Likert scale responses to all surveys and participant speaking time and researcher notes were recorded during each session. Qualitative data was analyzed by a single member of the team for themes which appeared in the formative study, as well as novel commonalities in behavior or themes emerging from the feedback. In this study, we embrace the approach of Yin et al. and focus on a small group of participants, mimicking real world conditions, and a longitudinal period of time [94]. A smaller group of participants more closely resembles real-world VC sessions and allows researchers to adequately analyze the rich qualitative data resulting from a longer term study. The focus was on capturing richness of user interaction and their perceptions.

7.5.3. Task

Before each session began, we asked for a volunteer to host. The host was responsible for facilitating session operation without interference from researchers. We only received one volunteer to host per session. After the host had been chosen, our research team remained with cameras and microphones off, only appearing at the end of each session to disseminate questionnaires and schedule the next session. Each
session featured a different task which lasted approximately 60 minutes. Tasks were chosen based upon their common reporting as VC tasks in our surveys, as well as those present in existing literature. Participants were given written documentation of the tools operation prior to the first session, as well as a tutorial walkthrough by our researchers at the beginning of each session.

- **Day 1**: Icebreaker task - participants were asked to prepare 1-3 slides introducing themselves and something they do for fun. Participants then discussed an ice-breaker question selected by the host from a list of common ice-breaker questions [19].

- **Day 2**: Essay Summary task - participants are given a short (1-3 page) essay to summarize (5 minutes) and lead a short (10 minute) discussion. The essays were taken from the same text used in the formative mock-classroom study discussed above.

- **Day 3**: Brainstorming co-design task - participants are given a design challenge and asked to work together to brainstorm solutions. This challenge was “imagine you have to explain to an alien how a bus works. How would you do this?” Whether to use external tools such as Whiteboard or Mirro were left to the discretion of the host, as well as how to delegate using these tools in collaboration with the other members of their session. Our host chose to use Whiteboard throughout the duration of the entire session.

Following each session, participants were given a brief questionnaire containing short answer and Likert Scale questions detailing their experience with each of the tools contained in Erato, and comparing these experiences to other VC contexts in which they had previously participated. At the conclusion of the final day, participants were also given another questionnaire comprising a System Usability Scale
Results Accessible Telepresence (SUS) as well as qualitative questions detailing their overall impression of the study and Erato. Following this, each participant engaged in a 45 minute semi-structured interview with a researcher, during which time they discussed their answers to the survey and elaborated on their experience. All exit interviews were conducted by the same researcher.

Section 7.6

Results

Participants found the system to be overall easy to learn, easy to use, and that they would use Erato frequently if it is available (See Figure 7.4). Likert ratings on a scale of 1 to 5 with 1 meaning “not at all” and 5 meaning “very much so” were comparable between participants who are d/DHH and non-d/DHH alike.

![Figure 7.4: 5 Point Likert Scale Response From Participants Who Are d/DHH to the Question “The Prototype Tool Made Me Feel More Confident While Presenting”. Results Reported are on a Scale of 1 to 5 with 1 meaning “not at all” and 5 meaning “very much so”.](image)

We observed that our 3 participants with d/DHH who participated in all 3 ses-
sessions felt more confident in their speaking ability with each subsequent session, which they attributed to using our tool. The 2 participants who joined for a single session were asked the same question, and we found that their result was within the standard deviation of our 3 participants who participated in all 3 sessions’ median score reported during the first session. This suggests that participant confidence may not be directly tied to specific tasks and seems to improve over time (See Figure 7.5). We probed which tool elements potentially resulted in speaking confidence and report these results below.

Figure 7.5: Average and standard deviations for Systems Usability Scale Results Reported by Participants in our case study. Results are on a scale of 1 to 5 with 1 meaning “not at all” and 5 meaning “very much so”.

Participants suggested a total of 9 accommodation requests across all three sessions, 5 of which were approved. Approved accommodation requests were maintained
across all future sessions. 3 accommodation requests were integrated into the etiquette guide during the first session, 1 during the second, and 1 during the third. All but one accommodation request was evaluated by the host, and some were omitted due to their commonality or overlap with other requests. Some accommodations which were requested were discussed orally before-hand by the group as the etiquette guideline pertained to a specific social tension arising organically during the session.

7.6.1. Evaluation of Individual Tools

We collected Likert scale data for each tool from all participants in all 3 sessions and found that scores were consistent and favorable across all 3 sessions during which that tool was used. This suggests that experiences and usability of individual tools were ecologically consistent regardless of task. We unpack the qualitative data and uses of these tools below. We present our results by grouping our tools according to their relevance to our design goals outlined in Section 3.

**Etiquette Guide (G1).** Both d/DHH and non-d/DHH participants indicated that this tool was the most useful of the tools supported by Erato: “Knowing everyone’s accommodations and being able to make mine known was very helpful and empowering” (P6). One of the key features of the tool is the anonymity provided by the interface, which ameliorates the pressure point of embarrassment we documented in our formative observational study: “It’s a great tool, especially for someone who may not be as comfortable asking for accommodations or doesn’t want to interrupt or draw attention to themselves to ask for adjustments” (P8). All meeting hosts found the review accommodation request feature easy to use, and appreciated having an opportunity to familiarize themselves with accommodation etiquette prior to facilitating the session: “It’s a lead by example sort of process. I know ahead of time what to do and can make sure I do what’s right” (P6). While providing the host monitoring
Table 7.3: Summary of accommodation requests made during the Erato case study.

<table>
<thead>
<tr>
<th>Accommodation Request</th>
<th>Session No.</th>
<th>No. Repeat Requests</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Please look directly at camera while speaking so I can read your lips”</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>“Please mute microphone when not speaking”</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>“Limit background noise in your environment if possible”</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>“Please say your name before speaking”</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>“Please say ‘over’ when done speaking”</td>
<td>3</td>
<td>0</td>
</tr>
</tbody>
</table>

The privileges of accommodation requests was intended to prevent potentially trolling or disruptive behavior on behalf of bad actors, none of the participants behaved this way. Interestingly, the bell button next to each approved accommodation request which sends a reminder of the associated etiquette to all participants in the session was seldom used. We unpack the social dynamics resulting from this tool below in Section 6.2. Participants differed on preferences for component location in the Erato tool window. Our initial design placed the notification center at the top of the window, assuming that this would be an intuitive place to consolidate feedback at-a-glance. However, all of our participants with d/DHH expressed a wish for the etiquette guide to be placed on top, citing that this was far more useful and should be visually prioritized. “I’d like to see the notifications area and the accessibility reminders swapped as that would be a better line of sight for those who also use captions” (P8).

**Speaker Assistive Tools (G2).** Most non-d/DHH participants monitored the notification center while presenting, while d/DHH participants monitored individual elements, specifically the speech speed and human volume buttons. Participants reported that they did so to overcome variable audio quality in individual speakers’ microphones. All participants who are d/DHH indicated that the speech speed button was their preferred tool for both presenting and watching other presents because
7.6 Results

adjusting speech speed had a noticeable impact on the quality of the automated captions. Furthermore, participants found the speech and volume tools to be reassuring they were presenting effectively, and felt more confident in their oral presentation quality as a result. Complications arose during more discussion-based activities where it was unclear who the volume and speech speed feedback was intended to address. Participants involved in all three sessions, however, reported that they adapted to watching the tool more closely when they were speaking, and assumed any notifications regarding speech speed and volume were directed at them if they appeared while they were speaking or within a small buffer of time before or after they spoke.

Host Specific Tools (G3). Hosts had access to additional tools not visible to other session participants. It is common for VC platforms to have a designated session host equipped with specific privileges and tools not afforded other session members to ensure that session activity can be monitored. Since all of our hosts were d/DHH, they did not require a reminder to turn on the captions since captions are a tool they would normally use. Hosts found that being able to quickly view data regarding the participation of everyone in the session was the most helpful for creating an inclusive environment for participants who are d/DHH. “Ensuring students are all participating, as well as making sure people (esp of different genders, races, etc.) in meetings are all being given a chance to be heard.” (P7). All participants with prior experience teach, in particular, remarked on the usefulness and necessity of this tool when adapting their classrooms to VC platforms. “You can’t always see what everyone is doing or keep track of who hasn’t spoken... seeing [this data] would keep students from slipping between the cracks” (P3). While the “limit text chat” and “are people texting or speaking” pie chart were both used at least once, we can not conclude how useful for devising intervention strategies these tools are for session hosts.
7.6.2. Emergent Socio-Technical Behaviors and Dynamics

Similar to behaviors documented in prior work involving reminder systems, we also observed that participants regarded the etiquette guide once at the beginning of each session, and only returned to review the etiquette guide when someone else was speaking to monitor their own behavior regarding accommodation requests [50]. Hosts similarly regarded the incoming accommodation requests at the beginning of the session, and wouldn’t integrate new accommodation requests submitted during the session task. As mentioned above, 1 accommodation request was not evaluated by the host since it was submitted close to the end of session 3, and was not seen by the session host. Interestingly, the “remind participants of etiquette” feature was only used twice across all three sessions. One explanation for this could be that new behaviors were incorporated into group dynamics quickly, and reminders were only necessary when large violations emerged, usually around single individuals. While new accommodation requests were introduced each session, participants easily incorporated these behavior requests into their social ecosystem and maintained these behavior changes across all future sessions. “It’s not an issue of malice, but of ignorance. No one is actively trying to make our lives as d/DHH people harder. They simply don’t know” (P7) This result suggests that creating awareness of accessibility needs in socio-technical spaces could produce substantial behavior changes, resulting in a more inclusive social ecosystem.

We also probed participants for social responsibility assessments to understand who in the group they assumed would be responsible for monitoring the etiquette guide and using the “remind participants of etiquette” feature. Two consensus emerged from this probe: one group suggested that monitoring these behaviors was the responsibility of those who submitted the request while the other group assumed it was the duty of the session host. “It just felt potentially rude and judgemental,
especially if the slip-up was tiny... I wouldn’t want someone to feel judged for messing up” (P4) Both of these groups agreed, however, that the main reason they didn’t use the “remind participants of etiquette” feature is because behavior changes and accommodation requests were readily adhered to by the group across sessions.

While the accommodation request tool alleviated experiences of social awkwardness by facilitating direct dialogue between participants with d/DHH, the host, and other session members, hosts experienced dilemmas around effectively intervening based on the data reflecting individual session participation. Each of our hosts mentioned that they noticed certain individuals who are d/DHH hadn’t participated in the discussion, but were not comfortable directly calling on them to participate. We probed all participants on their comfortable being called-on and found that most actively wanted to participate but felt they “couldn’t get in... the conversation was incredibly active and I didn’t want to be rude and interrupt” (P10). This highlights a tension between the mental model of host and participant desires for intervention provided data around individual participation.

7.6.3. Design Insights

While the results of our preliminary study suggest some benefit to all the tools comprising Erato, 3 tools in particular produced the most insightful findings. We thus present these three features below as design suggestions to inform the future development of VC platforms to be more inclusive and accessible.

**Anonymous Accommodation Request and Communication.** We demonstrated that anonymous accommodation requests not only alter behavior by bringing awareness of participants to accessible VC etiquette, but that these behavior changes remain consistent between sessions with different tasks and require little additional effort to maintain on behalf of hosts and participants with disabilities. It is our rec-
ommendation that these tools for anonymous accommodation request be integrated into existing VC platforms to further streamline the accessible and inclusive social interactions they afford, as well as address other accessibility and discriminatory concerns.

As noted in prior work, the embarrassment of making accommodation requests is mitigated in larger groups where anonymity of the person making the request is shielded by the size of the group [228]. However, this may not remain true in groups with only a single person with accessibility needs. While making accommodation requests is an anonymous process, the session participants' anonymity is removed if they are the only person with accessibility needs in the group. While our groups were large enough to provide some anonymity, we noted strategies emerging that could protect the identity of the person with accessibility needs if they were the only participant with accessibility needs in the session. We noted in our case study that both d/DHH and non-d/DHH participants made accommodation requests. On day 2 and day 3, non-d/DHH participants made repeat accommodation requests the noted on the first day. In addition, “please mute your microphone” was an accommodation request made by both d/DHH and non-d/DHH participants simultaneously. The instructor as well made accommodation requests on behalf of students.

**Gender Equality and Accessibility Beyond d/DHH.** We initially designed our tool to focus on ameliorating the accessibility limitations of speakers with d/DHH with a long-term vision of adapting these tools to assist and support people with other disabilities. Surprisingly, p3 brought to our attention that one benefit of visualizing individual participation was the potential for attenuating disparity and discrimination based on gender. “I wish this was required on all platforms because anytime women speak for more than 30% of any meeting discussion, men think that women are dominating the conversation” (P3). We probed this idea with other participants
who identified as women and found that all had similar experiences of feeling discriminated against when speaking during VC sessions. “It happens all the time [in the physical spaces] and you just get used to it or don’t notice it... having cold data makes it hard to dismiss and inexcusable.” (P3). For this reason, we suggest incorporating this tool into future design of VC platforms to highlight and potentially address the discrimination experienced by minorities when participating in VC spaces. Such a tool is helpful beyond accessibility for people who are d/DHH. “For accessibility, diversity, and feminism, we must adopt this tool into widespread use” (P3).

**Human to Human Feedback.** While our Speech Speed Meter and Volume Meter used automated feedback evident in other speech feedback tools, the frequent attenuation to the Human Volume Button and Speak Faster/Slower Buttons indicates that fully automated feedback is not enough. “I would use this immediately. I really wish I had this for church sermons. I attend church online and I would love to be able to discreetly tell the speaker to slow down so I can read their lips. They talk so fast I get lost. Even the captions get scrambled.” (P6). While automated tools for providing real-time feedback on participant volume and speed are a helpful baseline for producing more confident speakers, the specific needs of individuals observing the speaker vary. Providing a means of direct human-to-human feedback such as the Human Volume Button and Speak Faster/Slower Button ensure that the individual and subjective needs of the audience are attenuated. Supporting the individual needs of participants could broaden the inclusivity of VC platforms.
While the results of our preliminary study suggest some benefit to all the tools comprising Erato, 3 tools in particular produced the most insightful findings. We thus present these three features below as design suggestions to inform the future development of VC platforms to be more inclusive and accessible.

7.7.1. Anonymous Accommodation Request and Communication

We demonstrated that anonymous accommodation requests not only alter behavior by bringing awareness of participants to accessible VC etiquette, but that these behavior changes remain consistent between sessions with different tasks and require little additional effort to maintain on behalf of hosts and participants with disabilities. It is our recommendation that these tools for anonymous accommodation request be integrated into existing VC platforms to further streamline the accessible and inclusive social interactions they afford, as well as address other accessibility and discriminatory concerns.

7.7.2. Gender Equality and Accessibility Beyond d/DHH

We initially designed our tool to focus on ameliorating the accessibility limitations of speakers with d/DHH with a long-term vision of adapting these tools to assist and support people with other disabilities. Surprisingly, P3 brought to our attention that one benefit of visualizing individual participation was the potential for attenuating disparity and discrimination based on gender. “I wish this was required on all platforms because anytime women speak for more than 30% of any meeting discussion, men think that women are dominating the conversation” (P3). We probed this idea with other participants who identified as women and found that all had simi-
lar experiences of feeling discriminated against when speaking during VC sessions. “It happens all the time [in the physical spaces] and you just get used to it or don’t notice it... having cold data makes it hard to dismiss and inexcusable.” (P5). For this reason, we suggest incorporating this tool into future design of VC platforms to highlight and potentially address the discrimination experienced by minorities when participating in VC spaces. Such a tool is helpful beyond accessibility for people who are d/DHH. “For accessibility, diversity, and feminism, we must adopt this tool into widespread use” (P3).

7.7.3. Human to Human Feedback

While our Speech Speed Meter and Volume Meter used automated feedback evident in other speech feedback tools, the frequent attenuation to the Human Volume Button and Speak Faster/Slower Buttons indicates that fully automated feedback is not enough. “I would use this immediately. I really wish I had this for church sermons. I attend church online and I would love to be able to discreetly tell the speaker to slow down so I can read their lips. They talk so fast I get lost. Even the captions get scrambled.” (P6). While automated tools for providing real-time feedback on participant volume and speed are a helpful baseline for producing more confident speakers, the specific needs of individuals observing the speaker vary. Providing a means of direct human-to-human feedback such as the Human Volume Button and Speak Faster/Slower Button ensure that the individual and subjective needs of the audience are attenuated. Supporting the individual needs of participants will further broaden the inclusivity and accessibility of VC platforms.
Limitations and Future Work

This work explored the usability and effectiveness of a suite of VC tools to support speakers who are DHH, provide valuable data to session hosts regarding participation, and broaden the inclusivity of VC platforms by encouraging inclusive social etiquette. While our initial results show the promise of how incorporating such tools can improve the accessibility and inclusivity of VC sessions across 3 common tasks, we uncovered several opportunities for future work and investigation into how these platforms can be made more equitable.

7.8.1. Mitigating Bad Actors and Potential System Conflicts

Our study made the assumption that all participants would act in good faith when interacting with the system, and not abuse the tools or intentionally stress the system. While some design choices were made to directly mitigate potential issues (e.g. requiring the session host to approve accommodation requests) other opportunities for abuse are evident in the system. The human volume and speech speed buttons, for example, could be repeatedly pressed by a bad actor to intentionally frustrate the speaker. One way to prevent these potential issues would be to limit how frequently a user could press these buttons, possibly locking these features if persistent use is deemed abusive. Another approach could be to give additional controls to the session host for locking the tools of specific users, similar to a session hosts ability to mute microphones for participants. Still, even if acting in good faith, disagreements on ideal volume levels or speech speed may emerge from interacting with the system, and these struggles may prove disruptive for the participants. One potential solution to mitigating this problem is devise heuristics that detect potential disagreements, and provide individual suggestions for users based on their system state such
as "please increase your volume" if the user’s volume is turned down low, or "please follow along with this transcript" if the disagreement involves speech speed. While these disagreements did not emerge during our study, examining solutions to these conflicts would be an ample avenue for future work.

7.8.2. Embarrassment and Plurality
We also noted in our discussion above to the limitations of anonymity afforded by the accommodation request feature. In groups where only one person has a known disability, the system’s ability to minimize the potential embarrassment or discomfort evoked by making accommodation requests might be diminished. While we noted several emergent strategies that may mitigate these concerns, future work will explore how a greater plurality of representative abilities may affect the experience of users. While Erato focused on supporting participants who are DHH, future work will explore how other abilities can be supported using a similar system. Supporting a wider set of system supported abilities could help further mitigate the embarrassment or discomfort evoked by making accommodation requests by increasing the likelihoods of a plurality of represented abilities, and thus decreasing the likelihood that a single individual would be making requests. Additional work into eliminating embarrassment altogether remains the subject of future investigations.

7.8.3. Meeting Session Size and Scope
While we found compelling evidence that integrating tools comprising Erato into existing VC systems increases the confidence of speakers who are DHH as well as broadens the inclusivity and accessibility of VC platforms, these findings are confined to small discussion based groups. Many education and professional use-cases require large VC group meetings, and how the tools present in Erato would scale to such sessions is outside the scope of this initial investigation. Our formative study, for
example, documented tensions regarding the cognitive load incurred by a participant with DHH switching from reading captions to reading chat occurring in the text field. This phenomenon was not evident in our case study, and we do not have sufficient evidence to suggest that our tool in its current form resolved this tension. Future work will investigate how the tools comprising Erato scale to large group VC contexts, and how the scale of these gatherings affect the behaviors documented in our initial investigations.

### 7.8.4. Extension to ASL Signers

This work focused on users who are DHH that may not know ASL or have access to an interpreter. While considerable prior research exists exploring how to incorporate ASL signing into VC presentations, it remains an open problem. The method proposed in this work could be extended to incorporate the needs of ASL signers, as they could benefit from our approach. One avenue for future research in this domain could be to extend our experiment to incorporate signers and determine their usability by employing the ASL-SUS [97]. The intricacies and needs of this user-group are significant and nuanced, thus necessitating their own study, which we intend as future work.

### 7.8.5. Session Host Tools

Our tool suite was designed primarily to support DHH speakers, and we prototyped several new features for session hosts contributing to this goal. Session host tools were informed by prior literature as well as qualitative feedback form an experienced VC instructor, and additional feedback was provided by 3 session participants using Erato’s session host tools in our case study. While the initial feedback on these tools is promising, the results are limited by the small number of participants contributing feedback to the tools design. Future work will explicitly focus on improving tools
for session hosts who are DHH to support their capacity to effectively facilitate VC session tasks.

7.8.6. Personalizing Feedback for Discussions
As demonstrated in our user study, individuating the intended recipient of specific real-time feedback becomes difficult during discussions involving quick exchanges of speakers. This produces difficulties in not only understanding who individual pieces of feedback are the intended audience, but also which feedback is meant to be attenuated by the current speaker. One possibility would be to leverage the captions transcript as a source of individuation since current captioning systems are already able to distinguish between individual participants. Clicking a button on one of Erato’s tools and dragging to the name of a specific user intended to receive the feedback triggered by that button could ensure that the right feedback is attenuated by the right person, even in rapid exchanges. Examining methods for individuating this sort of feedback is the topic of future work.

7.8.7. Communal Agglomeration of Accessibility Etiquette Taxonomy
One benefit of Erato’s accommodation request system is that it maintains a library of previously approved accommodation requests, and preserves them across sessions. Similar to the approach proposed in prior work [50], sufficient dissemination of Erato could lead to two beneficiary outcomes. First, broader knowledge and awareness of accessible etiquette guidelines could be brought to the attention of diverse target end-users of VC platforms. This alone could greatly improve the accessibility of VC sessions and support long-term behavior change to make socio-technical dynamics more inclusive. Second, prolonged use of Erato by diverse users could over-time agglomerate a taxonomy of accessible etiquette guidelines, formalizing and communally aggregating behaviors which make VC platforms more accessible. While prior lit-
erature has documented some etiquette that could improve VC communication, our proposed approach could over-time create a large scale collection of accessible etiquette [150]. Monitoring and creating a consensus on these guidelines presents an interesting set of challenges which would be an ample topic for further discovery.

Section 7.9

Conclusion

While current VC systems are more accessible for various use-cases than physical environments for DHH users, they still mimic several forms of oppressive and discriminatory behavior evident in physical spaces. Addressing these concerns requires systems which can agglomerate and disseminate information regarding inclusive etiquette when participating in social activities mediated by socio-technical systems. We presented Erato, a tool suite comprised several existing tools evident in VC platforms as well as several new tools developed specifically to address the needs expressed by participants with DHH in our formative study. While all of these tools provide a modicum of assistance in improving the accessibility of VC platforms and supporting people with DHH when speaking during VC activities, 3 tools in particular show great promise. 1) A system enabling direct and anonymous communication of accessibility accommodation requests, monitored by the session host, and disseminated to all session participants, 2) A data-driven tool to allow hosts to monitor individual participation of session members as well as quickly understand who most recently spoke, and 3) A method for complementing automated speech speed and spoken volume that allows human-to-human communication of subject speech speed and volume needs. Our formative and case studies demonstrate that these three tools broaden the inclusivity and accessibility of VC sessions for 3 common tasks conducted on VC platforms. Furthermore, qualitative feedback from participants suggests that the
applicability of ameliorating discrimination experienced by minorities in VC social spaces extends beyond people who are DHH to people with other disabilities as well women and gender minorities. For these reasons and others mentioned in our study results, we encourage the consideration and adoption of these tools into the design of future VC platforms.
Appendix A

Idle-Time Analysis for Calliope System

In order to understand the idle time for each operation (mutation, interpretation of mesh editing, and generation) we first performed a series of each task sequentially. For each of these operations, we collected the time taken by each task. Since Calliope encourages parallel design iteration, it’s necessary to understand the idle wait times of operations in parallel. To evaluate this, we performed 2, 3, and 4 operations simultaneously and recorded the idle time. We did not evaluate generation operations in this way because the increase in latency was negligible.
Figure A.1: Average computational idle time for single tasks.
Figure A.2: Average computational idle time for parallel tasks.
Appendix B

Clio System Operations Reference

Below we detail the operations referenced throughout Chapter 6. We will first discuss the operations that we do support. Note that some operations were consolidated into generalized operations as discussed in the results of our formative user study:

**Previous Slide:** presenter is able to cycle to the previous slide in their prepared media. This became the more general operation “Previous” to allow the presenter to cycle through media other than slides like visual media.

**Next Slide:** presenter is able to cycle to the next slide in their prepared media. This became the more general operation “Next” to allow the presenter to cycle through media other than slides like visual media.

**Activate/Trigger:** presenter is able to begin specific behaviors using a trigger-action approach to mapping.

**Select:** presenter is able to select items from a menu of virtual objects on the screen.

**Close Menu:** presenter closes/collapses the chosen menu.

**Open Menu:** presenter opens/expands the chosen menu.

**Zoom 2D:** presenter is able to enlarge or shrink a 2D virtual object such as an image or text.
Laser Pointer: presenter is able to indicate various sections of virtual objects using a small dot which they control, similar using a laser pointer to indicate areas of interest on slides.

Expand Collection: presenter is able to expand a collection of objects away from each other to make them easier to view and manipulate.

Group Objects: presenter is able to cluster or group objects together to clean up the viewspace and make the collection easier to move.

Highlight: presenter is able to draw transparent colored annotations over a selected area to draw attention to it.

Annotate Object: presenter is able to draw on an object and have the resulting drawing track with that virtual object.

Annotation Air: presenter is able to annotate the viewspace itself.

Dismiss: presenter is able to remove virtual objects from the viewspace.

Conjure: presenter is able to make virtual objects appear within the viewspace.

Arrange: presenter is able to move, sort, and manipulate virtual objects around the viewspace.

***

The following operations were pantomimed during the second phase of our formative study but were not implemented:

Push/Pull Content from Chat: presenter is able to drag web links, text, or visual media into the viewspace or push media from the viewspace into the chatbox.

Poll/Quiz: presenter is able to generate and present a quiz for the audience to complete. Results can be displayed and manipulated on screen.

Add Shape: presenter is able to select from a variety of shape primitives such as circles, arrows, and others to appear onscreen.

Screen Grab: presenter is able to capture a section of the screen in an image.
and then use this image in their viewspace like any other visual media.

**Create Virtual Copy:** presenter is able to make virtual copies of virtual objects onscreen.

**Tangible Proxy:** presenter is able to use physical objects in their environment as targets for mapping images, animations, and behaviors. For example, a presenter may use index cards from their physical environment and map images to them. Thus, when they hold the index cards up to the screen, it appears as if they are holding the mapped image.

**Make Transparent:** presenter is able to lower the opacity of a virtual object.

**Rotate (3D):** presenter is able to rotate 3D assets in the viewspace.

**Zoom 3D:** presenter is able to enlarge or shrink a 3D virtual object.
Appendix C

Clio User Study Questionnaires

Below we include the questionnaires used in our study to collect our qualitative and Likert results.

Section C.1

Demographics

Multimodal Direct Manipulation in Telepresence Systems: Challenges and Opportunities Questionnaire 1: Demographics

The purpose of this questionnaire is to gather demographic information and prior experience with virtual presentations as part of a larger study. “Virtual presentations” includes any variety of computer communication platforms including Zoom meetings with slides, YouTube videos, or social media live-streaming.

1. Would you like to be contacted about future studies related to novel interfaces?
2. What is your participant number? Ask our research team if you do not know
3. What is your age?
4. What is your gender identity?
5. What is your occupation? Do you have any disabilities that you’d like to disclose or request for accommodation?
If you answered “yes” above, how might we accommodate you in this study?

Experience with Virtual Presentations

“Virtual presentations” includes any variety of computer communication platforms including Zoom meetings with slides, YouTube videos, or social media live-streaming.

6. Do you have prior experience either watching virtual presentations? (virtual presentations include slide shows such as Power Point on telepresence software such as Zoom, video lectures on platforms such as YouTube, or similar media)

7. If so, think of one recent example. Could you describe how the presentation was delivered? What software, media, etc.

Do you have prior experience either giving virtual presentations? (virtual presentations include slide shows such as Power Point on telepresence software such as Zoom, video lectures on platforms such as YouTube, or similar media)

8. If so, think of one recent example. Could you describe how the presentation was delivered? What software, media, etc.

9. What are the strengths of this presentation style?

10. What are the weaknesses of this presentation style?

11. Have you ever watched a virtual presentation that was somewhat improvised? If so, please explain

12. Have you ever given a virtual presentation that was somewhat improvised? If so, please explain

13. Do you have any experience using or witnessing presentations involving gesture recognition, voice activation, or computer-aided visual augmentation? If so, please explain your experience Would you be comfortable using non-mouse and keyboard driven interactions for presentations? Please explain your answer.

Virtual Presentations—Likert Scale
Please rate the following on a scale of 1 to 5 with 1 meaning “not at all” and 5 meaning “very much so”. If you have no experience with the questions material or are unsure, please leave the question blank.

“Virtual presentations” includes any variety of computer communication platforms including Zoom meetings with slides, YouTube videos, or social media live-streaming.

14. It is time consuming to prepare virtual presentations (“Virtual presentations” includes any variety of computer communication platforms including Zoom meetings with slides, YouTube videos, or social media live-streaming.)

15. It is difficult to prepare virtual presentations

16. I prefer watching virtual presentations with more images than text

17. I prefer delivering virtual presentations with more images than text

18. I am generally engaged when watching virtual presentations

19. I am more engaged watching virtual presentations in which I can see the speaker

20. When giving virtual presentations, it is easy for me to answer questions using my prepared content (slides etc.)

21. When giving virtual presentations, it is easy for me to answer questions midway through the presentation, or encourage interactivity with the audience

22. I feel compelled to ask questions or interact with the speaker when viewing a virtual presentation

23. I often interact or often see highly interactive virtual presentations

24. I like giving virtual presentations using traditional tools

25. I like watching virtual presentations using traditional tools

Closing thoughts

26. Is there anything further you would like to tell us about virtual presentations
or any of the topics mentioned in the questions above?

Section C.2

Audience Post-Performance

Multimodal Direct Manipulation in Telepresence Systems: Challenges and Opportunities

Questionnaire 3: Audience Post-Performance

The purpose of this questionnaire is to gather feedback regarding the performance of semi-extemporaneous presentations using a prototype creativity support tool.

Audience Impressions of Presentation

1. Please give an overall impression of the presentations you saw today
2. Were you engaged during the presentations? What factors contributed to your engagement?
3. Please comment on what you liked about this presentation style?
4. Please comment on what you did not like about this presentation style?
5. Please comment on how seeing the presenter co-located with their media affected the presentation
6. Were there certain interaction techniques you wanted to see used more? (interaction techniques include gesture, speech, mouse, etc.)
7. Were there certain interaction techniques you wanted to see used less?
8. Would you rather watch a long (30 minute) presentation using traditional virtual presentation tools (such as powerpoint on Zoom) or the prototype? Why?
9. What use-cases could you see this style of presentation being useful for?
10. Could you imagine ways in which this prototype could be useful for live presentations? Why or why not?
11. Were some presentations better than others? If so, what elements made some presentations better than others?
12. Were any interaction techniques overwhelming or corny? If so, which?

13. Do you think some interaction techniques would be better suited for some presentation contexts than others? If so, please explain.

14. Would you want to give a presentation using this prototype? Why or why not?

Likert Scale Questions: Please rate the following on a scale of 1 to 5 with 1 meaning “not at all” and 5 meaning “very much so”. If you have no experience with the questions material or are unsure, please leave the question blank.

15. I enjoyed watching presentations using the prototype

16. Did you notice any of the following interaction techniques being used?
Mouse Mid-air hand gestures Speech command recognition Tablet Keyboard

17. Did you notice any of the following operations being used?
arranging images text display on-screen drawing pan/zoom

18. If you noticed mid-air gesture interactions, were they engaging and enjoyable?

19. If you noticed voice interactions (e.g. commanding elements on the screen using spoken voice) were they engaging and enjoyable?

20. If you noticed on-screen drawing and annotation, were enjoyable and engaging?

21. If you noticed the use of text display, was it enjoyable and engaging?

22. I was generally engaged during virtual presentations using the prototype

23. I was more engaged with the virtual presentation using the prototype partially because I could see the speaker better than traditional virtual presentations

24. I felt compelled to ask questions or interact with the speaker when viewing the virtual presentation using the prototype

25. I would be more interested in interacting with a speaker giving a virtual presentation with the prototype

26. I liked watching virtual presentations with the prototype
27. I would rather watch a virtual presentation using the prototype than traditional presentation tools (such as slides on Zoom).

28. Please rank your preference for seeing the following interactions being used during the virtual presentations with 1 indicating “most favorite interaction” and 5 indicating “least favorite interaction”. If you didn’t notice one of these interactions, please select N/A.

   Gestures Mouse Keyboard Voice Command Tablet

29. Please rank your preference for seeing the following operations being used during the virtual presentations with 1 indicating “most favorite interaction” and 8 indicating “least favorite interaction”. If you didn’t notice one of these interactions, please select N/A.

   Arrange items Pan/Zoom Draw Text Display Conjure Dismiss Next Previous

30. Is there anything further you would like to tell us about virtual presentations or any of the topics mentioned in the questions above?

---

Section C.3

Speaker Post-Presentation

*Multimodal Direct Manipulation in Telepresence Systems: Challenges and Opportunities* Questionnaire 4: Presenter Post-presentation

The purpose of this questionnaire is to gather feedback regarding the performance of semi-extemporaneous presentations using a prototype creativity support tool.

Experience with Prototype (Performance Phase)

1. Please explain your overall impression presenting with the prototype

2. Do you feel it was helpful to prepare in a similar environment to that in which you presented?

3. How would you compare this experience with presenting a traditional virtual
4. Did the presentation proceed as you expected?
5. How did you manage unexpected incidents during the presentation?
6. Were there any interaction methods you wish you’d used instead of the ones you chose?
7. Which of the following interaction methods did you try during the rehearsal phase?
   - Gesture
   - Mouse
   - Keyboard
   - Speech
   - Tablet
8. Which of the following interaction methods did you use during your presentation?
   - Gesture
   - Mouse
   - Keyboard
   - Speech
   - Tablet
9. If you used gesture, did you find it enjoyable, useful, and intuitive? Why or why not?
10. If you used voice, did you find it enjoyable, useful, and intuitive? Why or why not?
11. If you used tablet, did you find it enjoyable, useful, and intuitive? Why or why not?
12. If you used on-screen drawing and annotation, did you find it enjoyable, useful, and intuitive? Why or why not?
13. If you used text display, did you find it enjoyable, useful, and intuitive? Why or why not?
14. Please rank your preference for using the following interactions during the virtual presentation with 1 indicating “most favorite interaction” and 5 indicating “least favorite interaction”. If you didn’t use one of these interactions, please select N/A.
   - Gesture
   - Speech
   - Mouse
   - Tablet
   - Keyboard
15. What are the strengths of this presentation style?
16. What are the weaknesses of this presentation style?
17. How did the presentation feel different using the prototype than traditional presentation tools?
18. How did it feel to be co-located with your presentation media (images, etc.)?
19. How would you have prepared differently if asked to present again?
20. Can you think of use-cases in which this tool would be helpful?

Likert Scale Questions: Please rate the following on a scale of 1 to 5 with 1 meaning “not at all” and 5 meaning “very much so”. If you have no experience with the questions material or are unsure, please leave the question blank.

21. It was easy giving virtual presentations using the prototype
22. I would use this prototype to prepare virtual presentations in the future
23. I trusted the system more because I prepared my presentation in the same environment in which it was delivered
24. I was skeptical of various interaction techniques (gesture, voice, etc) prior to presenting with the tool
25. I am more trusting of these interaction methods after presenting
26. The rehearsal process built trust between me and the tool
27. It was easy for me to answer questions mid-way through the presentation, or encourage interactivity with the audience using the prototype
28. I would prefer to present future virtual presentations using this tool than traditional presentation tools
29. I would like to see virtual presentations given with this tool
30. I improvised during my presentation more than anticipated
31. It was intuitive to give virtual presentations using the prototype
32. My presentation proceeded in a strict sequential order in which it is to be
Section C.4

Study 1: Formative Study Follow-Up

Multimodal Direct Manipulation in Telepresence Systems: Challenges and Opportunities Questionnaire 6: Formative Study Speaker Presentations

The purpose of this questionnaire is to gather feedback from presenters who we interviewed during our formative study. Participants will be asked to use our tool to present their proposed use-case. Participants will use our prototype to present their originally proposed use-case instead of pantomiming as they did in the formative study.

Experience with Prototype (Performance Phase)
1. Please explain your overall impression presenting with the prototype
2. Do you feel it was helpful to prepare in a similar environment to that in which you presented?
3. How would you compare this experience with presenting a traditional virtual presentation?
4. Did the presentation proceed as you expected?
5. How did you manage unexpected incidents during the presentation?
6. Were there any interaction methods you wish you’d used instead of the ones you chose?
7. What are the strengths of this presentation style?
8. What are the weaknesses of this presentation style?
9. How did the presentation feel different using the prototype than traditional presentation tools?
10. Were your needs met for presenting your original use-case?
11. How would you have prepared differently if asked to present again?
12. Can you think of other contexts in which this tool would be helpful?
13. What alterations would you suggest to make this tool better suited for your use case?
14. Were there any interaction techniques you were skeptical to use?
15. How would you say the rehearsal process affected your decision of what interaction methods to use?
16. Overall, was the prototype what you expected? If not, explain

Likert Scale Questions: Please rate the following on a scale of 1 to 5 with 1 meaning “not at all” and 5 meaning “very much so”. If you have no experience with the questions material or are unsure, please leave the question blank.

17. It was easy preparing and giving virtual presentations using the prototype
18. I would use this prototype to prepare virtual presentations in the future
19. It was intuitive to prepare and give my presentation using the prototype
20. I was skeptical of various interaction techniques prior to presenting with the tool
21. I am more trusting of these interaction methods after presenting
22. It was easy for me to answer questions using my prepared content with the prototype
23. It was easy for me to answer questions mid-way through the presentation, or encourage interactivity with the audience using the prototype
24. I would prefer to present future virtual presentations using this tool than traditional presentation tools.

25. I would like to see virtual presentations given with this tool.

26. I improvised during my presentation more than anticipated.

27. My presentation was better planned than I anticipated during rehearsal.

28. My presentation proceeded in a strict sequential order in which it is to be presented.

29. I was able to encourage interactivity during my presentation.

30. I liked using this prototype presentation tool.

31. My presentation closely resembled my pantomime during the previous phase of the study.

32. Is there anything further you would like to tell us about virtual presentations or any of the topics mentioned in the questions above?

---

**Section C.5**

**Commentator Post-Presentations**

*Multimodal Direct Manipulation in Telepresence Systems: Challenges and Opportunities* Questionnaire 5: Video Feedback

The purpose of this questionnaire is to gather feedback on a series of videos depicting an audience and a presenter interacting using a prototype virtual interaction tool.

1. What is your age?

2. What is your gender identity?

3. What is your occupation?

4. Do you have any disabilities that you’d like to disclose or request for accommodation?
5. Do you have prior experience either giving or watching virtual presentations? (virtual presentations include slide shows such as Power Point on telepresence software such as teams, video lectures on platforms such as YouTube, or similar media)

6. If so, could you describe how the presentation was delivered? What software, media, etc. Please be as specific as possible.

7. How would you compare the prototype presented in the video against your previous experience with virtual presentations?

8. Does the audience seem engaged?

9. How improvised are these presentations?

10. Would you rather present a talk using this prototype or traditional presentation tools? Why?

11. Would you rather watch a talk using this prototype or traditional presentation tools? Why?

12. Can you think of contexts where this prototype would be useful? Explain

13. While the presentations you watched during this study were short, do you think this style of presentation would work well for a longer presentation? Why or why not?

14. Any other observations or comments about the audience and presenter interaction from the video?

Likert Scale Questions: Please rate the following on a scale of 1 to 5 with 1 meaning “not at all” and 5 meaning “very much so”. If you have no experience with the questions material or are unsure, please leave the question blank.

15. The audience is engaged while watching virtual presentations using the prototype

16. The speaker is well prepared

17. The presentation is more engaging than traditional presentations
18. Is there anything further you would like to tell us about virtual presentations or any of the topics mentioned in the questions above?
Appendix D

Questionnaires from Clio

Validation Studies

To help with replication and future comparisons, we have included our surveys and questionnaires from our studies below.

Section D.1

Accessible Remote Video Conferencing Survey

This survey was used for collecting general responses to experiences using remote video conferencing software.

1. What is your age?
2. What is your gender identity?
3. What is your occupation?
4. Are you Deaf or Hard of Hearing?
5. How would you describe any disabilities you may have?
6. Do you have experience with remote video conferencing software such as Zoom or Teams?
7. What are some examples of times you have used video conferencing systems in the past? Please list as many as you can.

8. Describe some of the barriers or pressure points you experience during remote video conferencing sessions?

9. What are the best features of remote video conferencing environments? (e.g. Taking meetings from anywhere, less background noise, etc.) Do you use any accessibility tools during in-person lectures, meetings, or presentations?

10. If so, what are they and are they provided by you, your employer, or someone else?

11. Where do you typically take your remote video conferences?

12. What sort of difficulties do you face when giving presentations in remote conferencing environments?

13. Please compare your experience giving presentations or talking in remote environments versus physical environments.

14. I am able to customize remote video conferencing environments to suit my needs.

15. How well did the remote environment meet your accessibility needs without further customization?

16. I feel confident presenting or giving a lecture in remote video conferencing environments

17. I feel confident talking or socializing in remote video conferencing environments

18. I feel confident brainstorming in remote video conferences

19. I feel confident presenting in physical environments

20. I find it easier to present in remote video conferencing environments than physical environments
21. Please explain your answer to the previous question in detail I am able to communicate effectively in remote video conferencing environments.

22. I am able to communicate effectively in in-person presentation environments.

23. I am able to focus effectively during remote video conferencing sessions.

24. I am able to focus effectively during in-person presentation sessions.

25. I am comfortable communicating with meeting hosts to request accommodations for remote video conferences.

26. I am comfortable communicating with meeting hosts to request accommodations for in-person meetings or conferences.

27. Would you be willing to participate in a follow-up interview to this questionnaire? All follow-up interview participants will be compensated for their participation in the interviews scheduled at their convenience.

Section D.2

Mock Classroom Study Questionnaire

This questionnaire was given after each presenter during the mock classroom formative study.

1. Please enter your participant ID

2. Please enter the discussion section number (1-6) meaning which speaker just finished speaking?

3. What elements of the classroom interface do you find helpful for supporting and sustaining your engagement?

4. What elements of the classroom interface create tension or pressure points which makes remaining engaged in the classroom discussion or presentation difficult?

5. Please indicate from 1 to 5 with 1 meaning “not at all” and 5 meaning “very much so” how engaged you felt during this section
6. Please indicate from 1 to 5 with 1 meaning "not at all" and 5 meaning “very much so” how much of the material from the presentation and discussion you feel you retained.

**Section D.3**

**Use-Case Study Questionnaire**

The following questionnaire was given during the use-case study described in Chapter 6

1. Did you present or participate in today’s session?
2. How do you feel you performed overall during your presentation?
3. Could you comment on your overall experience using the prototype tool?
4. Did you use the prototype tool while presenting? Which elements of the prototype tool did you use while presenting? (If you did not use the tool, just put “NA”)
5. Which elements of the prototype tool did you use while other participants were presenting? (If you did not use the tool, just put “NA”)
6. Could you please compare your experience using the prototype tool today versus previous experiences without the tool?
7. Please rank the following elements of the prototype in order from “most useful” to “least useful”. If you didn’t use a specific element of the prototype, select “NA”. [Accessibility Accommodation Request] [Etiquette Guide] [Report Etiquette Violation] [Speech Volume Meter] [Speak Louder/ Quieter Buttons] [Speech Speed Meter] [Speak Quicker/ Slower Buttons] [Headphone Playback] [Ensure Captions Toggle] [Limit Text Chat Toggle] [Who Spoke Last Dashboard] [Text vs Spoken Chat Pie Chart]
8. Please explain your above answer
8. Please rate the following features of the prototype on a scale from 1 to 5 where 1 indicates “not at all useful” and 5 indicates “very useful”

9. Accessibility Accommodation Request - at the bottom of the chrome extension where requests for accommodation could be entered to become part of the etiquette guide [Rating]

10. Please explain your rankings in the above question.

11. Etiquette Guide - The list of accommodation requests submitted by members of the session [Rating]

12. Please explain your rankings in the above question.

13. Etiquette Reminder Alert - The ability for session participants to remind members of the session to be mindful of accommodation requests listed in the etiquette guide by pressing the bell button. [Rating]

14. Please explain your rankings in the above question.

15. Speech Volume Meter - Volume meter at top of chrome extension that provides feedback to the speaker regarding their speech volume [Rating]

16. Please explain your rankings in the above question.

17. Speech Rate Meter - Meter at top of chrome extension that provides feedback to the speaker regarding their speech speed [Rating]

18. Please explain your rankings in the above question.

19. Speak Louder/ Quieter Buttons - Buttons beneath volume meter to allow participants to suggest that the person speaking should speak louder [Rating]

20. Please explain your rankings in the above question.

21. Speak Quicker/ Slower Buttons - Buttons beneath volume meter to allow participants to suggest that the person speaking should speak slower or quicker [Rating]

22. Please explain your rankings in the above question.

23. Headphone Playback - Toggle switch which plays audio back through head-
24. Please explain your rankings in the above question.

25. Ensure Captions Toggle - Toggle switch to remind the host and session participants to turn on captions [Rating]

26. Please explain your rankings in the above question.

27. Limit Text Chat Toggle - Toggle switch that reminds session participants to keep text chat to a minimum [Rating]

28. Please explain your rankings in the above question.

29. (Host) Who Spoke Last Dashboard - Lists all session participants indicating who spoke last and how much time each participant has spoken during the session [Rating]

30. Please explain your rankings in the above question.

31. (Host) Text vs Spoke Chat Pie Chart - Pie chart demonstrating how much chatting has occurred via text in the chat window versus spoken chat in the meeting room [Rating] Please explain your above answer I am confident I presented well today [Answer]

32. The prototype tool was helpful while I was speaking [Answer]

33. The prototype tool helped me be aware of the needs of other people in my session [Answer]

34. Other people presenting during my session spoke confidently, clearly and well [Answer]

35. I used the “report” feature to encourage other participants in the session to adhere to the etiquette guide [Answer]

36. The prototype tool was difficult to use [Answer]

37. The prototype tool was overly complicated [Answer]

38. I had difficulty learning to use the prototype tool [Answer]
39. I was overwhelmed by the number of notifications provided by the prototype tool [Answer]

40. The prototype tool was distracting [Answer]

41. The prototype tool made me feel more confident while presenting [Answer]

42. The prototype tool helped make the session more inclusive for everyone [Answer]

43. I would use the prototype tool in future video conferencing sessions. [Answer]

44. Is there anything else you would like to tell us about your experience using the tool today?
Appendix E

Full Erato Formative Study

Participant Table

Table E.1: Description of participants in formative studies. Fields noted as N/A indicate participant declined to answer this survey question. An abbreviated version of this table containing only participants included in the live study was described in Chapter 7.

<table>
<thead>
<tr>
<th>Participant Number</th>
<th>Self-Described Disability</th>
<th>Age</th>
<th>Gender Identity</th>
<th>Occupation</th>
<th>Study Participation Portion</th>
</tr>
</thead>
<tbody>
<tr>
<td>FP1</td>
<td>Progressive hearing loss</td>
<td>20</td>
<td>Cis-Man</td>
<td>Student</td>
<td>Survey, Interview, Live Study</td>
</tr>
<tr>
<td>FP2</td>
<td>ADHD</td>
<td>21</td>
<td>Demi-Male</td>
<td>Student</td>
<td>Survey, Interview, Live Study</td>
</tr>
</tbody>
</table>

Continued on next page
Table E.1 – continued from previous page

<table>
<thead>
<tr>
<th>Participant Number</th>
<th>Self-Described Disability</th>
<th>Age</th>
<th>Gender Identity</th>
<th>Occupation</th>
<th>Study Participation Portion</th>
</tr>
</thead>
<tbody>
<tr>
<td>FP2</td>
<td>ASD</td>
<td>26</td>
<td>Cis-Woman</td>
<td>Student</td>
<td>Survey, Interview, Live Study</td>
</tr>
<tr>
<td>FP4</td>
<td>None</td>
<td>N/A</td>
<td>Cis-Woman</td>
<td>University Instructor</td>
<td>Survey, Interview, Live Study</td>
</tr>
<tr>
<td>FP5</td>
<td>bilateral hard of hearing since birth, low vision</td>
<td>36</td>
<td>Cis-Man</td>
<td>Student</td>
<td>Survey, Interview, Live Study</td>
</tr>
<tr>
<td>FP6</td>
<td>None</td>
<td>19</td>
<td>Cis-Man</td>
<td>Student</td>
<td>Survey</td>
</tr>
<tr>
<td>FP7</td>
<td>None</td>
<td>30</td>
<td>Trans-Man</td>
<td>Graduate Student</td>
<td>Survey</td>
</tr>
<tr>
<td>FP8</td>
<td>partial vision, low vision</td>
<td>28</td>
<td>Cis-Man</td>
<td>Unemployed</td>
<td>Survey</td>
</tr>
<tr>
<td>FP9</td>
<td>None</td>
<td>19</td>
<td>Cis-Woman</td>
<td>Student</td>
<td>Survey</td>
</tr>
<tr>
<td>FP10</td>
<td>bilateral hearing loss</td>
<td>28</td>
<td>Cis-Woman</td>
<td>Administration</td>
<td>Survey, Interview</td>
</tr>
<tr>
<td>FP11</td>
<td>None</td>
<td>32</td>
<td>Cis-Man</td>
<td>Vocalist</td>
<td>Survey</td>
</tr>
<tr>
<td>FP12</td>
<td>None</td>
<td>20</td>
<td>Cis-Man</td>
<td>Student</td>
<td>Survey</td>
</tr>
<tr>
<td>FP13</td>
<td>None</td>
<td>29</td>
<td>Trans-Woman</td>
<td>Lecturer</td>
<td>Survey</td>
</tr>
<tr>
<td>FP14</td>
<td>None</td>
<td>28</td>
<td>Cis-Woman</td>
<td>Teacher</td>
<td>Survey</td>
</tr>
</tbody>
</table>

Continued on next page
### Table E.1 – continued from previous page

<table>
<thead>
<tr>
<th>Participant Number</th>
<th>Self-Described Disability</th>
<th>Age</th>
<th>Gender Identity</th>
<th>Occupation</th>
<th>Study Participation Portion</th>
</tr>
</thead>
<tbody>
<tr>
<td>FP15</td>
<td>legally blind</td>
<td>31</td>
<td>Cis-Man</td>
<td>Unemployed</td>
<td>Survey, Interview</td>
</tr>
<tr>
<td>FP16</td>
<td>ADHD</td>
<td>20</td>
<td>Cis-Man</td>
<td>Student</td>
<td>Survey, Interview</td>
</tr>
<tr>
<td>FP17</td>
<td>Hard of Hearing and ADHD</td>
<td>36</td>
<td>Cis-Man</td>
<td>Administration</td>
<td>Survey, Interview</td>
</tr>
<tr>
<td>FP18</td>
<td>ADHD and Dyslexia</td>
<td>26</td>
<td>Cis-Woman</td>
<td>Student</td>
<td>Survey, Interview</td>
</tr>
<tr>
<td>FP19</td>
<td>None</td>
<td>19</td>
<td>Cis-Woman</td>
<td>Student</td>
<td>Survey</td>
</tr>
<tr>
<td>FP20</td>
<td>ASD</td>
<td>27</td>
<td>Cis-Man</td>
<td>Analyst</td>
<td>Survey, Interview</td>
</tr>
<tr>
<td>FP21</td>
<td>None</td>
<td>19</td>
<td>Cis-Woman</td>
<td>Student</td>
<td>Survey</td>
</tr>
<tr>
<td>FP22</td>
<td>None</td>
<td>22</td>
<td>Cis-Man</td>
<td>Food Service</td>
<td>Survey</td>
</tr>
<tr>
<td>FP23</td>
<td>None</td>
<td>26</td>
<td>Cis-Man</td>
<td>UX Design</td>
<td>Survey</td>
</tr>
<tr>
<td>FP24</td>
<td>Hard of Hearing and ASD</td>
<td>44</td>
<td>Cis-Woman</td>
<td>Unemployed</td>
<td>Survey, Interview</td>
</tr>
<tr>
<td>FP25</td>
<td>deaf</td>
<td>23</td>
<td>Two Spirit</td>
<td>Student</td>
<td>Survey, Interview</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Participant Number</th>
<th>Self-Described Disability</th>
<th>Age</th>
<th>Gender Identity</th>
<th>Occupation</th>
<th>Study Participation Portion</th>
</tr>
</thead>
<tbody>
<tr>
<td>FP26</td>
<td>deaf and immuno-compromised by chronic heart condition</td>
<td>18</td>
<td>Cis-Woman</td>
<td>Student</td>
<td>Survey, Interview</td>
</tr>
</tbody>
</table>
Appendix F

Commercial VC Analysis

Here we collect results of the commercial VC platform analysis we describe in Chapter 7.

Table F.1: Benefits and pain points of Microsoft Teams as described by DHH/HoH survey and interview participants in Formative Study (See Chapter 7).

<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Make a written record of your meeting</td>
<td>Captioning is a little obscure to find</td>
</tr>
<tr>
<td>Captioning seemed more accurate than Meet and Zoom</td>
<td>Accessibility tools not on main menu bar (requires searching)</td>
</tr>
<tr>
<td>Can add interpreters to meetings</td>
<td>Mobile apps for Teams have limited capabilities</td>
</tr>
<tr>
<td>Captions can distinguish between speakers very well</td>
<td>Person sharing screen cannot see captions</td>
</tr>
<tr>
<td>Captions can catch when rapidly switching between speakers</td>
<td>No visual representation of speakers volume</td>
</tr>
<tr>
<td>Caption size is bigger Zoom</td>
<td>Unless you are a paying (subscription) Teams member, you do not get captions (financial inaccessibility)</td>
</tr>
<tr>
<td>Name of person listed in transcript and captions</td>
<td>Caption speed not adjustable</td>
</tr>
<tr>
<td>Caption size adjustable</td>
<td>Captions not visible in breakout rooms</td>
</tr>
</tbody>
</table>
Table F.2: Additional benefits and pain points of Zoom compared to Teams and Meet as described by DHH/HoH survey and interview participants in Formative Study (See Chapter 7).

<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Has captioning and captioning is a free service</td>
<td>Breakout rooms do not have captions</td>
</tr>
<tr>
<td>Pop up message notifies speaker that mic is muted</td>
<td>Transcripts not available for breakout rooms</td>
</tr>
<tr>
<td>Allows third-party captioning (which can be better than the built-in standard captioning) but these services often require additional fees</td>
<td>For HoH participants, it is difficult to tell whether the headphones aren’t working or people are speaking low</td>
</tr>
<tr>
<td>Supports having a live transcriber who types what is being said</td>
<td>Captions are very inaccurate (worse than both Meet and Teams)</td>
</tr>
<tr>
<td>Captions displayed while sharing screen (this feature is not evident in Meet nor Teams)</td>
<td>Captions disappear when using direct message feature on mobile devices</td>
</tr>
<tr>
<td>Create a live transcription</td>
<td>Live transcripts are slow to produce and very inaccurate</td>
</tr>
<tr>
<td>Visually displays volume of person speaking</td>
<td>Captions often confuse who is speaking and are unable to detect when the speaker switches rapidly</td>
</tr>
<tr>
<td>Direct messaging supported if enabled by meeting host</td>
<td>Captions display filler words</td>
</tr>
<tr>
<td>Font size of captions is adjustable</td>
<td>Captions do not adjust to screen size</td>
</tr>
<tr>
<td>Captions can detect and display a person speaking while sharing a video via screen sharing</td>
<td>When speaker shares a video via screen sharing, their audio is significantly reduced while other meeting participants can still speak at full volume</td>
</tr>
</tbody>
</table>
Table F.3: Benefits and pain points of Google Meet as described by DHH/HoH survey and interview participants in Formative Study (See Chapter 7).

<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Has captioning and switching between captioning languages is a free service</td>
<td>Non-English captions are slower and far more inaccurate than English captions</td>
</tr>
<tr>
<td>Text is bigger than Zoom</td>
<td>Captions are slightly inaccurate</td>
</tr>
<tr>
<td>Camera quality is better than Teams and Zoom</td>
<td>Lacks transcript capabilities</td>
</tr>
<tr>
<td>Can add Chrome extensions to extend capabilities of Meet</td>
<td>Caption speed cannot be adjusted</td>
</tr>
<tr>
<td>Lists the name of person speaking in captions</td>
<td>Participants cannot be directly messaged</td>
</tr>
<tr>
<td>Captions can detect quick switching between speakers</td>
<td>Unless you are a paying (subscription) Teams member, you do not get captions (financial inaccessibility)</td>
</tr>
<tr>
<td>Adjusts screen to fit the captions onscreen</td>
<td>The person sharing the screen can’t see the captions</td>
</tr>
<tr>
<td>Captions also available in breakout rooms</td>
<td>No visual indication of spoken volume</td>
</tr>
<tr>
<td>Captions can detect and display when two participants speak at the same time</td>
<td>Captions are only displayed on the tab of the Meet session and are not displayed on other tabs nor while sharing screen</td>
</tr>
<tr>
<td>Captions are robust to background noise and music</td>
<td>Captions not displayed when using whiteboard function or other tools and chrome extensions</td>
</tr>
<tr>
<td>Captions can detect and display audio from video shared via screen sharing</td>
<td>Does not caption laughter nor ambient sounds</td>
</tr>
</tbody>
</table>
Bibliography


on Computers and Accessibility (New York, NY, USA), ASSETS ’16, Association for Computing Machinery, 2016, p. 25–32.


[37] Hsiang-Ting Chen, Tovi Grossman, Li-Yi Wei, Ryan M. Schmidt, Björn Hartmann, George Fitzmaurice, and Maneesh Agrawala, *History assisted view au-


[43] John Joon Young Chung, Hijung Valentina Shin, Haijun Xia, Li-yi Wei, and Rubaiat Habib Kazi, Beyond show of hands: Engaging viewers via expressive and scalable visual communication in live streaming, Proceedings of the 2021


[47] Josh Davis, PokerFace Mask: Exploring Augmenting Masks with Captions through an Interactive, Mixed-Reality Prototype, January 2022 (eng), Accepted: 2021-12-24T17:47:44Z.


[50] ———, Circuitstyle: A system for peripherally reinforcing best practices in hardware computing, Proceedings of the 32nd Annual ACM Symposium on User
Interface Software and Technology (New York, NY, USA), UIST ’19, Association for Computing Machinery, 2019, p. 109–120.


[56] Jiachun Du, Panos Markopoulos, Qi Wang, Marina Toeters, and Ting Gong, ShapeTex: Implementing Shape-Changing Structures in Fabric for Wearable Ac-


on Human Computer Interaction (New York, NY, USA), Interacci& #xf3;n ’16, Association for Computing Machinery, September 2016, pp. 1–2.


[85] Jan Gugenheimer, Katrin Plaumann, Florian Schaub, Patrizia Di Campli San Vito, Saskia Duck, Melanie Rabus, and Enrico Rukzio, The impact of assis-


[153] Susan M. Mather and M. Diane Clark, *an issue of learning the effect of visual split attention in classes for*.


[172] Martez Mott, John Tang, Shaun Kane, Ed Cutrell, and Meredith Ringel Morris, “I just went into it assuming that I wouldn’t be able to have the full experience”: *Understanding the Accessibility of Virtual Reality for People with Limited Mobility*, (2020) (en-US).


[177] Gary Ng, Joon Gi Shin, Alexander Plopski, Christian Sandor, and Daniel Saakes, *Situated Game Level Editing in Augmented Reality*, Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Inter-
action (New York, NY, USA), TEI ’18, Association for Computing Machinery, March 2018, pp. 409–418.


[189] Fabio Paternò and Federico Gianmarino, *Authoring interfaces with combined use of graphics and voice for both stationary and mobile devices*, Proceedings of


[208] Matthew Seita, Sarah Andrew, and Matt Huenerfauth, *Deaf and hard-of-hearing users’ preferences for hearing speakers’ behavior during technology-mediated in-person and remote conversations*, Proceedings of the 18th Inter-
national Web for All Conference (New York, NY, USA), W4A ’21, Association for Computing Machinery, 2021.


York, NY, USA), ASSETS ’15, Association for Computing Machinery, 2015, p. 231–238.


IEEE International Symposium on Mixed and Augmented Reality (ISMAR),


[253] Ariel Weingarten, Ben Lafreniere, George Fitzmaurice, and Tovi Grossman, *Dreamrooms: Prototyping rooms in collaboration with a generative process*, Pro-


[264] Xiaoyi Zhang, Tracy Tran, Yuqian Sun, Ian Culhane, Shobhit Jain, James Fogaerty, and Jennifer Mankoff, *Interactiles: 3d printed tactile interfaces to enhance mobile touchscreen accessibility*, Proceedings of the 20th International ACM
SIGACCESS Conference on Computers and Accessibility, ASSETS ’18, ACM, October 2018.