2019-12-20

Evaluating User Preferences for Augmented Reality Interactions for the Internet of Things

Chopra, Shreya

Chopra, S. (2019). Evaluating User Preferences for Augmented Reality Interactions for the Internet of Things (Unpublished master's thesis). University of Calgary, Calgary, AB.

<http://hdl.handle.net/1880/111455>

Downloaded from PRISM Repository, University of Calgary

UNIVERSITY OF CALGARY

Evaluating User Preferences for Augmented Reality Interactions for the Internet of Things

by

Shreya Chopra

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE
DEGREE OF MASTER OF SCIENCE

GRADUATE PROGRAM IN COMPUTER SCIENCE

CALGARY, ALBERTA

DECEMBER, 2019

© Shreya Chopra 2019

UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled “Evaluating User Preferences for Augmented Reality Interactions for the Internet of Things” submitted by Shreya Chopra in partial fulfillment of the requirements of the degree of Master of Science.

Supervisor, Dr. Frank Oliver Maurer, Department of Computer Science

Dr. Mario Costa Sousa, Department of Computer Science

Dr. Nelson Wong, Department of Computer Science

Abstract

This thesis investigates how users want to control IoT devices in their homes with headset AR. Gestural and voice controls are both suitable methods of input for headset AR. However, there is a lack of end-user input in the design of such gestures and voice commands: especially in comparative terms. An elicitation study is performed with 16 participants to gather their preferred voice commands and gestures for a set of referents. The contribution is an analysis of 784 inputs (392 gestures and 392 voice), a resulting gesture set and command types, comparative method preferences, a novel method to analyze voice command input (called **voice command pattern template**), as well as design recommendations based on observations and interviews. These recommendations serve as guides for future designers and implementors of such voice commands and gestures.

Keywords: Augmented Reality, Internet of Things, Gestural Interaction, Voice Commands, User-Centered Design, Human Computer Interaction, Elicitation Study

Acknowledgements

- To Dr. Frank Maurer, thank you for being a constant source of belief. Your confidence in me allows me to believe that I can keep achieving more and more. It is truly said that one should look for a good boss rather than a good position or workplace, and I was lucky enough to find everything. I have really grown under your watchful, yet freedom-granting eye. You are an inspirational leader who is always willing to help everyone, and you know how to pull the strengths of everyone in a team together. Not only have I learned a lot about computer science from you, I have also learned countless soft skills that I will implement for the rest of my career. You are really a great supervisor- thank you!
- To Claudia Maurer, thank you for handing me that first piece of paper and encouraging me to apply for my first research job after I graduated. At that point, I was really a wide-eyed student trying to maneuver my next step, and you showed me the way. Thank you for seeing something in me. That was the job where I gained confidence to pursue research and teach data science. I will always cherish the design, networking, data science, and leadership skills that I learned from you. I will also always remember your plants & chocolate!
- To Dr. Wesley Willett, thank you for hiring me on my first research job after I graduated. I learned so much about design and data science from you that I later tried my best to pass down to my TAs and students and into my own projects. Thank you for the leadership that we knew was always there, but that we never felt. Thank you for your invaluable support whenever I needed it. The opportunity that you gave me has truly pushed my horizons and I will be forever thankful for it!

- To Dr. Sheelagh Carpendale, thank you for hiring me on my first research job after I graduated. Other than data science, I also learned so much about art and design from you. Your way of working through challenging meetings and tasks with a relaxed vibe is something that passes onto the entire team and is something that I look up to. Thank you for having trust in me to deliver results- it always pushed me forward!

- To my ASE Lab Mentors:

Omar Addam- thank you for being the first person in the lab to help me transition in. I don't know how you managed to do it even though you were so busy with your own things. Your help on my course project enabled me with an endless amount of confidence and opportunity in a software environment that was totally new to me. I hope to pass forward the software skills and soft skills onto others like you did for me!

Dr. Teddy Seyed- thank you for offering your wise advice whenever I needed it. Even though you are extremely busy, I know you are there to offer advice when I need it. I look up to you for your creativity and malleability to be able to pick up new skills. Thank you for being such a good role model!

Lorans Alabood- thank you for always being literally around the corner so that I could bounce ideas off you and get insight regarding my topic from someone more experienced. It was really nice to have you there since you were the only other person working in IoT/ Smart technology!

Dr. Tiago Silva da Silva- thank you for being available to answer questions about HCI and research whenever I needed it. Your calm ability to work on multiple things at once is truly inspirational!

- To my fellow past and present ASE Lab members: Zahra Aminolroaya, Cooper Davies, Nathan Douglas, Bijan Farhoudi, Kun Feng, Hadiya Firdaus, Micheal Friesen, Patrick Gharib, Bryson Lawton, Rachel Lee, Alec McAllister, Christian Roatis, Angela Rout, Hannah Sloan, Austin Sullivan, Bobbi Tang, Tarishi Upadhyay, Teresa Van, Steven Vi, Dianna Yim: You are some of the toughest, most inspirational people that I know. Each of you has inspired me in your own unique ways. Perhaps our uniqueness is what makes us such a great team. I wish you all the best, and I know you will do well for yourselves!
- To the ASE interns that I have mentored on other lab projects: Ana DuCristea, Alexander Greco, Rakshit Patel, Nihal Potdar: You have awed me with your ability to ask countless questions and be able to discover things all at the same time. I have learned so much from working with you!
- To the students that I have TAed: You completed my master's journey. I discovered things about myself that I did not know existed. I have grown with you!
- To Edwin Chan and Qing (Lion) Chen- thank you for the work that you guys did before me. I have referred to each of your thesis countless times and felt grateful for them!
- To my examination committee members:

Dr. Nelson Wong- thank you for taking on the task of being on my committee. In the past, I have learned a lot from you about teaching from marking assignments to engaging with students: all with ease. Thank you also for your constant support. I will carry forward my learnings forever!

Dr. Mario Costa Sousa- Even though I never had the pleasure of working with you, I would like to thank you for taking on the task of being on my committee!

Dedication

To My Family.

“You raise me up, so I can stand on mountains

You raise me up to walk on stormy seas

I am strong when I am on your shoulders

You raise me up to more than I can be”

-Josh Groban

Table of Contents

Abstract.....	iii
Acknowledgements	iv
Dedication	vii
Table of Contents	viii
List of Figures and Illustrations	x
 Chapter 1: Introduction	 1
1.1 Augmented Reality.....	2
1.2 The Internet of Things.....	7
1.3 Home Usage Scenarios.....	10
1.4 Types of Major Controls: Gesture and Voice	11
1.5 Motivation.....	14
1.6 Research Goals	15
1.7 Research Questions.....	16
1.8 Research Methodology	18
1.9 Thesis Contribution	18
1.10 Thesis Structure	19
 Chapter 2: Background and Related Work	 20
2.1 Augmented Reality Technology.....	20
2.2 IoT Technology	23
2.3 AR & IoT Combined Research.....	26
2.4 Elicitation Studies	30
2.5 Agreement Rate Evaluations	34
2.6 Grounded Theory: Open Coding	38
 Chapter 3: User Elicitation Study for AR-IoT Controls.....	 39
3.1 Study Design.....	39
3.1.1 Pilot Study Participants	39
3.1.2 Participants	39
3.1.3 Apparatus.....	40
3.1.4 Scenarios.....	42
3.1.4.1 <i>Interacting with a Menu System</i>	43
3.1.4.2 <i>Environmental Control</i>	51
3.1.4.3 <i>Media Control</i>	55
3.1.4.4 <i>Following a Workflow</i>	59
3.1.5 Rounds	64
3.1.6 Procedure	65
3.2 Data Analysis.....	66
3.2.1 Observational Methodology & Classification of Input	67
3.2.1.1 <i>Gesture Set</i>	67
3.2.1.2 <i>Voice Commands Set</i>	73
3.2.2 User Preference.....	81
 Chapter 4: Study Observations, User Comments, and Design Implications	 83

4.1 Gesture Round Observations.....	83
4.2 Voice Round Observations.....	85
4.3 Overall Observations.....	86
4.4 User Comments.....	87
4.4.1 Interaction and Strategy	87
4.4.2 Preferences	89
Chapter 5: Limitations & Conclusion.....	92
5.1 Scope & Limitations	92
5.1.1 Scope	92
5.1.2 Limitations	92
5.2 Future Work & Conclusion	95
5.2.1 Future Work	95
5.2.2 Conclusion	98
References	100
Appendix A: Pre-Study Questionnaire	118
Appendix B: Part 1 (Gestures first script).....	119
Appendix B: Part 2 (Voice first script)	121
Appendix C: Results of Pre-Study Questionnaire	123
Appendix D: Raw Data.....	124
Appendix E: Gesture Analysis.....	132
Appendix F: Voice Command Analysis	134
Appendix G: Non-Abstracted (Detailed) Results Table	136
Appendix H: Storyboards	137

List of Figures and Illustrations

Figure 1: LG smart TV [Left], Phillips Hue lightbulb [Center], and Nest thermostat [Right] are all examples of devices that are “smart” or embedded with Internet capabilities. These are usually capable of being remotely controlled via apps or home control systems such as the Google Home. [Image sources: 70, 113, 123].	1
Figure 2: Commercial examples of current augmented reality. The Pokémon Go game app [Left] and the Ikea Place app [Right] both demonstrate that most commercial AR is based on touch screens such as phones and tablets [image sources: 69, 59].	2
Figure 3: An example of the emergence of headset-based AR in Research & Development. Thyssenkrupp prototypes what it would be like to offer elevator maintenance workers headset-based guidance [49].	3
Figure 4: Demonstrates an example of an AR scenario. There are three components in this scenario: the user (real person) wearing an AR headset, a physical TV, and a holographic remote control for the TV (which only the wearer of the headset can see). This example is used to demonstrate the 4 ways of being able to interact in augmented reality (as demonstrated by the following 4 figures).	5
Figure 5: Firstly, the user can manipulate physical objects as if they were not wearing the headset at all. Here, the user goes to the TV and turns it on (without a remote control). They can walk right through the holographic remote and directly manipulate the physical object.	5
Figure 6: Secondly, the user can manipulate holographic objects almost as if the physical world around them does not exist. Here, the user increases the size of the holographic remote control.	6
Figure 7: Thirdly, the user can manipulate physical objects to alter virtual objects. Here, as the user increases the volume on the physical TV, the volume level shows as increased on the holographic remote as well.	6
Figure 8: Fourthly, the user can manipulate virtual objects to alter physical objects. Here, when the user presses the virtual power button on the holographic remote control, the physical TV turns off.	7
Figure 9: Example of IoT/ smart devices that may co-exist within a network. These devices can be discovered on the network and can also be controlled via a control point if one exists [Image source: 98].	8
Figure 10: Current forms of IoT Controls. [Top] Phillips Hue phone app to control smart lights. [Bottom] Google Home and Amazon Echo voice control devices to control various smart devices [Image Sources: 120, 75].	9
Figure 11: Example of an IoT control from our study. This thesis defines IoT controls as the combination of the control method (gesture in this case) and the augmented reality	

interface (holographic clock here). This control is being used to interact with or manipulate the physical IoT device (light). The clock is used to set the light automation times. This is also an example of headset AR based IoT controls.....	10
Figure 12: In a generic household like the one pictured here, there are some devices that may be smart or IoT-enabled: TV, lights, fan, dishwasher. However, with the rapid increase of embedded computability, in the future, the table may be able to keep track of what one is eating or the sofa may track health stats. [Image source: 111]	11
Figure 13: Various methods of input. [Top Left] Skin-based input, [Top Center] touch screen input, [Top Right] physical button input, [Bottom Left] voice command input, [Bottom Center] eye-tracking or gaze input, [Bottom Right] gestural input. [Image sources: 45, 30, 76, 19, 126, 6]	12
Figure 14: An example of a home IoT scenario in which the user is controlling devices through a Microsoft HoloLens. The user can be seen performing a “bloom” gesture to access information, view media, and perhaps even control appliances in their home [73]..	13
Figure 15: Overview of user defined gestures for opting-in and opting-out from interactions with public displays from <i>Rodriguez et al.</i> ’s elicitation [103].....	14
Figure 16: The most frequently used/suggested voice and gesture commands to control an unmanned aerial vehicle from <i>Peshkova et al.</i> ’s elicitation [92]. This is a rare piece of work that is a comparative elicitation between voice and gestures with emphasis on both. The aim is to help fill this gap: especially in terms of augmented reality for IoT.....	15
Figure 17: Various AR headsets. [Top Left] Lenovo ThinkReality, [Top Right] Microsoft HoloLens, [Bottom Left] DreamWorld DreamGlass, and [Bottom Right] Meta’s Meta 2. We use the HoloLens. [Image Sources: 117, 121, 35, 64].....	21
Figure 18: Various AR glasses. These are less bulky in comparison to the headsets. However, they may also offer lesser capabilities. [Top Left] Epson Moverio BT-300, [Top Right] Magic Leap One, [Bottom] Google Glass Enterprise Edition. [Image Sources: 108, 36, 48].	22
Figure 19: Examples of touch screen based augmented reality. [Left] User interacts with the Houzz app on an iPhone via Apple’s ARKit. [Right] User interacts with an educational app on an Acer tablet via Google’s ARCore. [Image sources: 32, 1].....	23
Figure 20: Example of a platform that lets users program the way IoT devices behave. Image of the Reality Editor AR app developed by <i>Huen et al.</i> from MIT Fluid Interfaces lab. Here, the user programs his car’s thermostat to turn on when he leaves his office chair [50].	24
Figure 21: Examples of common IoT enabled home devices. [Top Left] Ecobee smart thermostat, [Top Center] iSmartAlarm security camera, [Top Right] Roomba smart vacuum, [Bottom Left] August smart lock, [Bottom Center] Bose smart speakers, [Bottom Right] Hotpoint smart washing machine. [Image sources: 29, 56, 31, 17]	25

Figure 22: Examples of emerging IoT enabled home devices. [Top Left] Click and Grow smart indoor garden, [Top Center] iHome smart plug, [Top Right] Anova Culinary Precision smart cooker, [Bottom Left] Velux smart blinds, [Bottom Center] Sobro smart coffee table, [Bottom Right] Husqvarna automower. [Image sources: 16, 29, 118, 112]	25
Figure 23: <i>Alce et al.</i> contribute three interaction models (floating icons, world in miniature, and floating menu). This is an image of their world in miniature model [4].....	29
Figure 24: <i>Mylonas et. al</i> use AR and IoT for educational purposes. These are sample photos from the educational lab kit and AR tool in practice [81].	29
Figure 25: <i>Hüttenrauch et al.</i> ’s table depicts the commands covered by the ETSI standard (European Telecommunications Standards Institute) [52].....	31
Figure 26: <i>Morris et al.</i> ’s table depicts the number of occurrences for the voice and gesture interactions that reached the maximum threshold of 3 or above. This makes up the user-elicited interactions set for interacting with a web-browser in a living room.....	33
Figure 27: <i>Chen et al.</i> ’s table depicts the gap for a valid calculation that accounts for multi-proposals by single participants in which the consensus amongst all participants’ proposal is 0. <i>Chen et al.</i> proposes a formula to fill this gap [26].	36
Figure 28: A quarter of the participants use headset-based AR 2-5 times a week while the remaining 12 participants have not used it at all or used it very rarely. No one has used it to control IoT devices in a non-work-related setting.	40
Figure 29: Example of the <i>physical</i> and <i>holographic</i> components as seen by the wearer of the headset. In this IoT home control scenario, the video display is the physical component while the video 1 thumbnail is the holographic (or virtual, graphical) component. The physical components can be seen by everyone while the holographic components can only be seen by the wearer of the headset.....	42
Figure 30: Connecting a Computer to a Print Queue Interface with the sequential steps. Initially, the holographic link button appears next to the physical computer.	44
Figure 31: Next, the user selects the link button: either via voice commands or gestures (as decided by the researcher).....	44
Figure 32: All dynamic menus expanded. Once the link button is selected, the option to either <i>Connect as Input</i> or <i>Connect as Output</i> appears. Once the user selects <i>Connect as Input</i> , the options to select <i>Print Queue</i> or <i>Music</i> appears. Once <i>Print Queue</i> is selected, the list of available printers appears. The user selects <i>Printer 3</i> . The interface then indicates that the computer is “Connected to Printer 3”. Lastly, the user is asked to collapse the entire menu so that only the initial link button remains.....	45
Figure 33: Setting Up a Lights Schedule Interface with the sequential steps. Initially, the holographic clock button appears next to the physical light. The user is looking up to see this.....	46

Figure 34: Next, the user selects the clock button: either via voice commands or gestures (as decided by the researcher).....	47
Figure 35: Once the clock is selected, the remainder of the menu will appear straight ahead so that the user does not have to keep craning their neck up. Once the options to pick either <i>Time</i> or <i>Day</i> appear, the user selects <i>Time</i> via voice or gesture.	47
Figure 36: Next, the options to select either <i>Start</i> or <i>End</i> time appear. The user selects <i>Start</i> . (Note that the small white circle follows the user's gaze, but that is not used as an input in our experiment.).....	48
Figure 37: Next, the user makes use of the respective <i>Hours</i> , <i>Minutes</i> , <i>Seconds</i> sliders to set the start time. Then, the user toggles to either AM or PM.....	48
Figure 38: Next, the user selects the <i>Done</i> button to indicate that they have finished entering the start time.....	49
Figure 39: Next, the corresponding menu to enter <i>end</i> time appears (and the <i>start</i> time menu collapses). The sequence for entering <i>end</i> time occurs in the same way.....	49
Figure 40: Next, the user selects <i>Day</i> . Then, the <i>time</i> menu collapses and the <i>day</i> menu will expand.	50
Figure 41: The user toggles on the days that they want the lights schedule to apply to (i.e. Monday). Next, they select <i>Done</i> to indicate that they have finished entering the days. Lastly, the user is asked to collapse the entire menu so that only the initial clock button remains.	50
Figure 42: Blinds control interface. A television display denotes a window where the paper with vertical lines is used to communicate the orientation of the blinds to the participants. The three buttons on the holographic component denote, respectively: rotate left, open/close window, rotate right.	51
Figure 43: The user holds down the <i>rotate right</i> button (and it turns black). The user was first asked to rotate blinds left, then rotate them to the right, then open them so that they are not covering the window any more.....	52
Figure 44: Thermostat Control Interface. Sequential steps as follows in the subsequent figures.	53
Figure 45: The user selects the <i>up</i> button to increase the temperature.	53
Figure 46: The user selects the <i>down</i> button to decrease the temperature.....	54
Figure 47: The user selects the <i>close</i> button to close the thermostat control.....	54

Figure 48: Speaker control interface. The speakers are at an unspecified location (or can be imagined as embedded on the ceiling). The user selects a holographic <i>audio thumbnail</i> to play it. Sequential steps as follows in the subsequent figures.	55
Figure 49: The <i>progress bar</i> replaces the audio thumbnail when that is selected. Here, the user selects the <i>pause button</i> to pause the song.	56
Figure 50: The <i>play button</i> is selected by the user to resume playing the song.	56
Figure 51: Video control interface. The user makes it so that <i>Video 1</i> will play on <i>Display 1</i> . There are two displays identified to the user (one being on their left and another being on their right). They are asked to make it so that the video will play on Display 1 when they eventually decide to play it.	57
Figure 52: Next, the user picks the <i>video thumbnail</i> so that the video will actually play.	58
Figure 53: The <i>progress bar</i> replaces the video thumbnail when that is selected. The user closes the video player so that it will stop playing.	58
Figure 54: Cooking bread dough interface with the sequential steps (in subsequent figures). This “kitchen counter” holds most ingredients that are needed. Initially, the user reads the first instruction, adds water, and selects the <i>next button</i> to go to the next step.	60
Figure 55: Next, the user reads the instruction, adds yeast, and selects the <i>next button</i> to go to the next step. (The <i>previous button</i> can be used to go to the previous step).	60
Figure 56: Next, the user reads the instruction, adds cooking oil, and selects the <i>next button</i> to go to the next step. In the same way, the user proceeds through the remaining steps until there are no more steps (<i>next button</i> is not available).	61
Figure 57: One of the steps require the user to go to a different counter to add flour (to simulate a kitchen scenario).	61
Figure 58: Fixing a boombox interface. Subsequent figures show sequential steps. Initially, the user reads the instructions, turns the boombox backwards, and selects the <i>next button</i> to go to the next step.	62
Figure 59: The user reads the instructions and finds the correct screwdriver. From here, they can decide to use the <i>next button</i> to go to the next step or the <i>previous button</i> to go to the previous step.	63
Figure 60: The user reads the instruction, places the antenna on the boombox, and selects the <i>next button</i> . In the same way, the user progresses through the instructions until the last step (when the next button is no longer available).	63
Figure 61: The tools that the user has to use to fix the boombox sit on the counter.	64

Figure 62: All possible 16 combinations of tasks. Half of the participants performed combinations 1-5-9-13 for round one and 4-8-12-16 for round two. The second half of the participants performed combinations 2-6-10-14 for round one and 3-7-11-15 for round two.	65
Figure 63: 12 unique gestures identified out of 392 gesture inputs. These gestures were sometimes performed with the entire arm stretched or with both hands, and these details are indicated in the raw data in Appendix D. A combination of gestures was also used sometimes. A gesture set was narrowed down from these gestures.	68
Figure 64: Atomic gestures often made up one gestural proposal. It was noted that drag was the most common factor amongst participants to perform this task.	69
Figure 65: Each participant’s bag of proposals. All proposals for each participant make up the bag for that participant. For example, participant 1’s bag is indicated with yellow highlighter: “air tap, air tap, air tap”. Participant 7’s bag is indicated with pink: “swipe, tap, tap”.	70
Figure 66: Each participant’s bag of proposals is narrowed down to a set.....	70
Figure 67: The final set is determined by combining all the participants’ sets.	70
Figure 68: The final set is numbered by the amount of people who ever proposed the gesture...	71
Figure 69: The gesture proposed by the most amount of people is declared and is supplemented with its agreement rate.	71
Figure 70: Overview of final gesture set. It should be noted that the “collapse menu” referent had a very low agreement rate for gestural input.	72
Figure 71: Voice command analysis table for picking a button referent. The final command and agreement rate emerge by separately analyzing both tasks and combining their results.	78
Figure 72: Voice command analysis table for going to next/previous step referent. This is a more complex example of analysis.	79
Figure 73: “Ok” is used at least 3 times in proposal 6, so it appears as a part of the emerging command. It is only used once in proposal 7, so it does not appear there.	79
Figure 74: “Ok” is not used as a part of emerging command for all proposals. However, “next” + “step” is a common factor throughout. Thus, “ok” is removed from the task’s emerging command.	79
Figure 75: “step” is used as a part of the cooking a recipe task’s emerging command. However, it is not used as a part of the fixing a boombox task’s emerging command. Nonetheless, “next” is a common factor in both. Thus, “step” is removed from the final command, and “next is considered as the final command.	80

Figure 76: Final Referent Table with the Voice Command Set and Gesture Set. A non-abstracted (more detailed version of this table is available in Appendix G.....	80
Figure 77: Users' Method of Preference.....	81
Figure 78: Reasons given by users for picking the preferred method	82

Chapter 1: Introduction

This thesis investigates user preferences for interacting with the Internet of Things (IoT) using augmented reality (AR) headsets as output devices and gesture or voice input. The Internet of Things is the connection of everyday objects that are embedded with the Internet, enabling them to send and receive data to each other [55]. This includes devices being able to generate data and output it. Examples of such devices may be: televisions, lights, and thermostats. Usually, if these devices are embedded with Internet capabilities, they can be controlled by phone apps (such as the Phillips Hue app to control smart lights) or voice-controlled home systems such as the Google Home. This thesis investigates what it would be like to control these IoT devices with augmented reality that uses a headset. In specific, the hope is to see how users *want* to interact using gestures and voice commands. Things that are looked for are: what kind of gestures and voice commands people want to use and whether they prefer one control method over another. A requirements elicitation study is contributed to help inform designers and implementers of such interactions with the users' perspective.

Figure 1: LG smart TV [Left], Phillips Hue lightbulb [Center], and Nest thermostat [Right] are all examples of devices that are “smart” or embedded with Internet capabilities. These are usually capable of being remotely controlled via apps or home control systems such as the Google Home. [Image sources: 70, 113, 123].

1.1 Augmented Reality

Immersive technology provides an environment that surrounds the user. Virtual reality (VR) and augmented reality (AR) are both examples of immersive technology, and this thesis focuses on augmented reality. AR is technology that superimposes a computer-generated image on a user's view of the real world, thus providing a composite view [9]. Recent commercial examples of such technology are Pokémon Go and the Ikea Place AR app. Pokémon Go is a game that was launched in 2016 that uses geographical information and displays holographic¹ characters in the real environment that the user must “catch”. The Ikea Place app displays holographic furniture in the real environment making it simple for the user to decide on potential furniture. As most other commercially-used AR, these apps are also based on touch screen technology such as smartphones and tablets.

Figure 2: Commercial examples of current augmented reality. The Pokémon Go game app [Left] and the Ikea Place app [Right] both demonstrate that most commercial AR is based on touch screens such as phones and tablets [image sources: 69, 59].

¹ “Holographic” is used as a shortcut for “virtual holograms”. “Holographic” usually refers to a 3-dimensional image.

However, headset-based AR technology is attempting to segue into commercial usage. Headset AR consists of wearing a glasses-like device on the head so that the physical environment and holographic components that are usually seen through a phone or tablet are now visible through the glasses. There are apps being released on the commercial front for headsets such as the Microsoft HoloLens. Although this is the case, the current price range of such headsets prevents it from being accessible to the mass public. Moreover, the bulkiness of such headsets paired with the small field of view that they currently offer deters public users from investing at the stage they are in right now. Nonetheless, headset companies strive to bridge these two gaps to eventually bring them to the hands of the users. Furthermore, headsets are at the forefront for AR researchers and industry, and ideal usage is at the stage of being investigated by those in Research & Development. One example is ThyssenKrupp's probe into elevator maintenance using AR [49].

Figure 3: An example of the emergence of headset-based AR in Research & Development. Thyssenkrupp prototypes what it would be like to offer elevator maintenance workers headset-based guidance [49].

Headset-based augmented reality is valued for some specific benefits that it offers. Firstly, the workspace/ real-estate that AR offers can be more seamlessly utilized via a headset simply because interaction is not limited to the screen space of the tablet/ phone. Also, holographic components are 3-dimensional, and users can even walk through them and all around them. This offers a very close-up view for cases in which holographic components need to be closely examined. Also, headset AR offers a level of privacy in which those around the user cannot easily see what the user is seeing. As the holographic components can only be seen with the headset on. As well, headsets allow for a hands-free approach since the headset does not have to be held up. Users can simply use their hands in thin air for gesture interaction or voice commands to prompt the system. Moreover, the user views both holographic components and the physical environment through the same perspective with the headset as opposed to touch-screen-based AR. For example, when using a tablet, the holographic components are only viewable and interactable through the screen of the tablet whereas with the headset, holographic components can be seen through the turn of one's head in the right direction- much like with physical objects.

In general terms, augmented reality offers 4 types of interactions (as seen in the images below). The merger of the physical and virtual world allows the user to manipulate: 1) physical objects as if they were not wearing the headset at all, 2) holographic /virtual objects almost as if the physical world around them does not exist, and 3) a combination of manipulating physical objects to alter virtual objects and 4) manipulating virtual objects to alter physical objects. Conclusively, there are multiple levels of controls for users of this technology.

Figure 4: Demonstrates an example of an AR scenario. There are three components in this scenario: the user (real person) wearing an AR headset, a physical TV, and a holographic remote control for the TV (which only the wearer of the headset can see). This example is used to demonstrate the 4 ways of being able to interact in augmented reality (as demonstrated by the following 4 figures).

Figure 5: Firstly, the user can manipulate physical objects as if they were not wearing the headset at all. Here, the user goes to the TV and turns it on (without a remote control). They can walk right through the holographic remote and directly manipulate the physical object.

Figure 6: Secondly, the user can manipulate holographic objects almost as if the physical world around them does not exist. Here, the user increases the size of the holographic remote control.

Figure 7: Thirdly, the user can manipulate physical objects to alter virtual objects. Here, as the user increases the volume on the physical TV, the volume level shows as increased on the holographic remote as well.

Figure 8: Fourthly, the user can manipulate virtual objects to alter physical objects. Here, when the user presses the virtual power button on the holographic remote control, the physical TV turns off.

1.2 The Internet of Things

The Internet of Things (IoT) is the ever-growing network of physical objects that feature an IP address for internet connectivity, and the communication that occurs between these objects and other Internet-enabled devices and systems [116]. These devices are inclusive of chips that enable them to generate, output, and receive data and be able to “talk” to each other. There are multiple types of IoT or “smart” capabilities that a device may have [86]. *Audio and Video* allows devices to access servers and render media. *Cloud* capability allows the device to store and access data on the Internet. *Device Management* means that the device has management capabilities such as troubleshooting, configuration, etc. *Home Automation* is specific to devices that are home-based. These include services like solar protection blinds, digital security cameras,

lighting controls, and HVAC (for thermostats). *IoT Management and Control* is the capability of a device to act as a bridge between the network and sensors & actuators. *Multiscreen* allows for displaying the same media simultaneously on multiple devices: such as a TV and a phone. *Networking* is a capability which is embedded in a device that acts as a gateway between the residential Local Area Network (LAN) and Wide Area Network (WAN), and through this, the device facilitates the smooth functioning of other IoT devices. *Printing* enables the functions of a smart printer (one that communicates over the network). *Remoting* allows home devices to communicate with remote devices. *Scanning* enables the functions of a smart scanner (just like a smart printer). *Telephony* enables management of incoming and outgoing phone calls and messages, etc. There is one more capability which can be referred to as *Basic Devices* which are low-cost entry points into the smart technology world. These contain very minimal interaction capabilities but can be discovered and minimally controlled through a network. These are examples of what major IoT capabilities are offered in today's smart devices.

Figure 9: Example of IoT/ smart devices that may co-exist within a network. These devices can be discovered on the network and can also be controlled via a control point if one exists [Image source: 98].

Many of these IoT capabilities and devices are seen in homes. However, these devices could be used in multiple scenarios: homes, offices, control rooms for industrial/ factory purposes. The focus of this thesis is on the home scenario. An IoT control is any medium used to interact with or manipulate IoT devices. This includes phone apps such as the Philips Hue app to control lights as well as home control systems such as the Google Home. For this thesis, IoT controls are defined as being the combination of the control method (voice and/or gesture) and interface (augmented reality holographic components) used to manipulate or interact with physical IoT devices.

Figure 10: Current forms of IoT Controls. [Top] Phillips Hue phone app to control smart lights. [Bottom] Google Home and Amazon Echo voice control devices to control various smart devices [Image Sources: 120, 75].

Figure 11: Example of an IoT control from our study. This thesis defines IoT controls as the combination of the control method (gesture in this case) and the augmented reality interface (holographic clock here). This control is being used to interact with or manipulate the physical IoT device (light). The clock is used to set the light automation times. This is also an example of headset AR based IoT controls.

1.3 Home Usage Scenarios

As of now, home devices that have IoT capabilities are things like lights, garages, thermostats, televisions, speakers, etc. As the practice of embedding computers into everyday devices increases, having a dinner table keeping track of what food the user eats or having a couch keep track of the user's health vitals is not far-fetched. The purpose of smart devices is to make everyday life easier for the user, and thus there is an increasing effort to make more devices that are IoT-based. Designers regard affordances (or what the environment offers) when creating these devices. Examples of things to keep mind from an AR perspective are the following. How do users want to expand a menu for a light that is high up? Should the expanded menu “come down” to

head level so that users avoid craning their neck? When is it more convenient to talk to the system rather than use gestures and vice versa? What happens when a user is controlling a device from far away? When does the user have their hands full and cannot interact with a device? Going off these ideas, this thesis probes on *how* users want to interact with such devices in home scenarios. More specifically, 8 home IoT tasks are developed, front-end prototypes are created, and users are asked to interact with gestural and voice command input.

Figure 12: In a generic household like the one pictured here, there are some devices that may be smart or IoT-enabled: TV, lights, fan, dishwasher. However, with the rapid increase of embedded computability, in the future, the table may be able to keep track of what one is eating or the sofa may track health stats. [Image source: 111]

1.4 Types of Major Controls: Gesture and Voice

Today's input technologies include touch screen input, non-screen-dependent gestures, voice commands, physical buttons and knobs, eye-tracking (gaze), input using the skin: to name a

few. The most commonly used are physical buttons and knobs (including computer keys), voice commands, and touch screen input. For the headset augmented reality that is used for this thesis, the most common inputs are gestures (made in thin air), voice commands, and gaze. Gaze is usually used analogous to a cursor on a computer. As mentioned above, this thesis focuses on gestures and voice commands.

Figure 13: Various methods of input. [Top Left] Skin-based input, [Top Center] touch screen input, [Top Right] physical button input, [Bottom Left] voice command input, [Bottom Center] eye-tracking or gaze input, [Bottom Right] gestural input. [Image sources: 45, 30, 76, 19, 126, 6]

Gestures are the most commonly used method of input for AR headsets. However, current hardware restricts users to a pre-defined set of gestures. The user makes use of their fingers and hand to “select”, “click”, or bring up a menu, etc. These gestures are usually determined by what is easy for the hardware to recognize. Examples of gestures are the “air tap” and “bloom” which are used for the Microsoft HoloLens.

The voice method is another method that comes secondary to the gestural input. When using the Microsoft HoloLens, it is especially used when an annotation appears over a button that enables users to read it to “click” on the button. Other than this, there are other commands that are recognized by the hardware. This method is also restricted with a pre-defined set of commands. In this thesis, the aim is to see what types of voice and gestures users *want* to use if there are no pre-defined restrictions.

Figure 14: An example of a home IoT scenario in which the user is controlling devices through a Microsoft HoloLens. The user can be seen performing a “bloom” gesture to access information, view media, and perhaps even control appliances in their home [73].

1.5 Motivation

Headset AR comes with methods of input that provide the user a pre-determined set of interactions. A literature review, which is detailed in chapter two, found that there is a gap in terms of eliciting end-user preferences for augmented reality interactions. There is some work done to elicit gestural interactions: not necessarily augmented reality-based. Some of these are *Chan et. al's* elicitation on single hand micro gestures [22] and *Wu et al.'s* elicitation on gesture interaction for immersive virtual reality shopping applications [124]. In addition, there is a significantly lesser amount of elicitation done in terms of voice commands. One such example is *Hüttenrauch et. al's* elicitation on commands for mobile devices and services [52]. There are only a few elicitations performed in terms of both gesture and voice. These include *Ghosh et. al's* elicitation on mobile voice assistants (with gesture as a secondary aspect) [40], *Rodriguez et. al's* elicitation on interactions with public displays (with voice as a secondary aspect) [103], and *Peshkova et. al's* elicitation on user-defined gestures and voice commands to control an unmanned aerial vehicle [92]. None of these comparative elicitations are based on augmented reality. Thus, this thesis aims to help fill this gap. More specifically, the hope is to provide insight into scenarios based on smart homes (IoT).

Figure 15: Overview of user defined gestures for opting-in and opting-out from interactions with public displays from *Rodriguez et al.'s* elicitation [103].

		<i>Up & Down</i>		<i>Left & Right</i>		<i>Rotate Left & Rotate Right</i>		<i>Forward & Backward</i>		<i>Takeoff & Land</i>	
Survey	Voice	up (81) upward (8) down (83) downward (4)		left (82) west (6) right (84) east (6)		turn left (30) rotate left (26) rotate counter clockwise (17) turn right (31) rotate right (25) rotate clockwise (17)		forward (58) go (13) straight (8) north (6) back/backward (42) reverse/south (6)		up (28) takeoff (21) lift/start (17) land (56) down (23)	
	Voice	up (88) higher (12) down (100)		left (80) turn left (12) right (92) turn right (8)		turn (48) turn left (12) turn (50) turn right (12)		forward (56) straight (32) back (76) backward (16)		up (42) start (32) land (40) stop (33) down (13)	
Experiment	Gestures	31 58		24 (left) and 26 (right)		35		23 35		-	
	Voice	7	9	8	10	10	9	7	7	9	15
Interview	Gestures	33 37		30 30		22		22 37		-	
	Gestures	9	10	9	9	8	9	9	9	2	3
	V+G	17	14	16	14	15	15	17	17	22	15

Figure 16: The most frequently used/suggested voice and gesture commands to control an unmanned aerial vehicle from *Peshkova et al.*'s elicitation [92]. This is a rare piece of work that is a comparative elicitation between voice and gestures with emphasis on both. The aim is to help fill this gap: especially in terms of augmented reality for IoT.

1.6 Research Goals

This thesis has four primary goals. The first is to inform designers and implementers about the current state of immersive IoT controls, especially in terms of augmented reality controls for IoT home scenarios. An extensive literature review is conducted to comprehend the current state

of augmented reality as well as IoT controls in terms of current hardware, application, and research. Research in gestures and voice commands in terms of AR and Non-AR is also extensively covered in the literature review. The second goal is to determine an elicitation method for involving users in the design of headset AR based IoT controls. More specifically, the aim is to determine *how* people like using immersive IoT controls. What are the gestures that people want to use? What are the voice commands that people want to use? The third goal is to provide a useful set of guidelines for the design of headset AR based IoT controls. These would ideally improve subsequent studies and implementations of immersive IoT controls. The aforementioned guidelines will be based on the results of the elicitation study. The fourth goal is to determine if users have a preference for voice control versus gesture control when navigating the controls.

1.7 Research Questions

To emphasize the relevance of this research and provide a concise contextual overview of the research space to potential designers of headset AR based IoT controls, the first question that arises is:

1. *What is the current state of research regarding immersive IoT controls (especially based on headset AR)?*

Answering this question will achieve the first research goal by informing audiences about the headset AR based IoT controls research space. Despite AR and IoT both being buzzwords that are valued as cutting-edge concepts in industry and research, there is not a lot of research work on a combination of these concepts.

2. *How can gestures and voice commands be elicited to trigger specific AR-IoT control tasks from users?*

Although there are no elicitation studies for headset AR based IoT controls, there are a range of elicitation studies involving similar methods to navigate other types of systems. These existing studies are valuable for their methods as well as findings. Also, utilizing a similar study method will aid us by providing useful data for evaluating the method itself. After determining the appropriate elicitation methodology, the study can be conducted to address the third research goal.

From the results of the study, the goal is to answer:

3. *What insights can be derived from the elicitation study results to make recommendations to future designers and implementers of headset AR-based IoT controls?*

The results of the study are analyzed both quantitatively and qualitatively. A gesture and voice command count is conducted and agreement rates are calculated to determine if there is a consensus. A gesture set as well as a voice command set is proposed- detailing the most popular gestures and voice commands for each referent. Themes and patterns are examined within the proposed gestures and voice commands. In addition, user comments are referenced to further any findings. These observations lead into determining the answer for the fourth research goal:

4. *Do users have a preference between voice control versus gesture control for headset AR based IoT controls?*

Just like the methodology for extracting answers for the third goal, this goal also consists of qualitative and quantitative analysis. A count of how many users prefer voice commands over gestures and vice versa is conducted. Furthermore, this is supplemented by their comments regarding each of the two methods.

Research questions #1 and #2 are addressed in Chapter 2 through an extensive literature review. Questions #3 and #4 are examined in Chapters 3 and 4 which discuss the results of the

elicitation study along with observational findings and user comments. Chapter 5 discusses the scope, limitations, future work, and conclusion.

1.8 Research Methodology

Wobbrock et al.'s requirements elicitation methodology [122] is used in which the *gulf of execution* [51] is removed so as to accommodate any action performed by the user as acceptable by the system. This ensures that the user is not limited by any limitations that a fully-implemented system may possess. In this case, that could have been the ability of the HoloLens to recognize certain gestures or accents when using voice commands. Creating a mock-up system where the front end is implemented but manipulated by the researcher as required by the user ensures freedom to the user. In this way, any gesture or voice command administered by the user is always successful. The study round of the elicitation method is further elaborated on in Chapter 2 and application of this methodology is discussed in Chapter 3.

1.9 Thesis Contribution

The contribution includes a study on headset AR based IoT controls especially in terms of gesture & voice commands. There is classification of a total of 784 elicited inputs in which 392 are elicited gestures and 392 are elicited voice commands (See **Appendix D** for the complete dataset). This is followed by the statistical analysis of the data using *chen et al.*'s revised agreement rate [26]. Furthermore, there is contribution of user preferences between the two methods of gesture control versus voice control. A novel method of analyzing voice commands is also contributed: *voice command pattern template*. This method is a major contribution since this kind of analysis has not been done before. Lastly, user comments are used to develop a set of design guidelines that offer qualitative insight into end-user thinking when designing gestures and voice commands that are meant to control IoT devices through headset AR: especially for a home

scenario. The scope of gesture and voice control is broad and offers countless variations, and this work is a preliminary effort to designing good gestures and voice commands to accommodate the affordances that users deal with and enhance their experience.

1.10 Thesis Structure

This chapter provides a concise background regarding the research for this thesis. The primary topic of headset AR based IoT controls is explained in this chapter. Headset AR, IoT, home usage scenarios, control types (gesture and voice), and motivation are explained here. The research goals, questions, methodology, and contributions are outlined.

Chapter 2: Background and Related Work – provides an overview of related research. The initial portion documents the existing research in AR, IoT and controlling IoT devices with headset AR. The latter portion reviews existing studies and methodologies for conducting gesture elicitation studies and voice elicitation studies.

Chapter 3: User Elicitation on Augmented Reality for IoT Controls– delivers the design of the elicitation study, the analysis techniques, and the quantitative results.

Chapter 4: Study Observations, User Comments, and Design Implications– combines the feedback generated from the post-task interviews with other qualitative analysis. Themes and motivations identified from the results aid in generating proposed design guidelines for controlling IoT devices with headset AR.

Chapter 5: Limitations & Conclusion– documents the depth and breadth of the scope of the study and discusses the challenges and restrictions derived from the limitations. It touches upon potential future work and confers the significance of the thesis contributions.

Chapter 2: Background and Related Work

In this chapter, we provide an overview of the existing research and current technology that is relevant to our work. We cover augmented reality and the Internet of Things. We also cover existing research in gestures and voice commands. We discuss the elicitation methods that are helpful to us.

2.1 Augmented Reality Technology

Augmented reality is currently available commercially in terms of headsets, tablets, and phones. Headsets are used mostly for research given the price, weight, and small field of view. However, they offer a hands-free approach to augmented reality given that users do not have to hold a device, but rather wear it on their head. There are various manufacturers of headsets. The Microsoft HoloLens, which we worked with, provides interaction through gaze, gesture, voice, and a hand-held clicker if required [43, 44]. The recent HoloLens 2 provides additional input capabilities and it is sleeker and possesses a more powerful processor than its predecessor [18]. There is an effort to integrate augmented reality capabilities into virtual reality headsets which usually do not allow the user to see the physical environment around them [41]. Facebook's recent Oculus Quest lets the user intermittently view a low-resolution view of the physical environment to avoid dangers (such as running into something) [104]. Input for virtual reality-based headsets uses physical controllers in each hand, and Oculus also offers voice input [46]. There are also augmented reality glasses (pictured in figure 17) which are less bulky than other headsets, but each manufacturer offers different capabilities [48, 10]. For example, the Google Glass which was deemed unsuccessful commercially, has some use cases for industry [54]. That can be controlled via a touchpad on the glasses as well as with voice commands [48]. Also, the Magic Leap One glasses require the user to sling on their shoulder a tiny computer that runs the glasses [91], and

the user can input using Headpose (the rotation of your head), Eye Gaze (where you look at and what you are focusing at), Gesture (hand gestures using your real hands), Voice Commands, Controller (using the bundled controller and future compatible accessories), Mobile app, Keyboard [8]. The Epson Moverio products offer a range of inputs depending on the hardware: their headset has more evolved voice and gesture capabilities than their glasses [44, 37]. The Lenovo ThinkReality AR headset offers voice control, a hand-held controller, and gaze control [71, 117]. While the ThinkReality is known as a HoloLens competitor- especially since it is more lightweight, the user must carry another component that runs the ThinkReality while the HoloLens is a standalone device [71]. The Meta and DreamGlass headsets both function only when tethered to a computer, and the Meta offers gestural input while DreamGlass also incorporates voice input [90]. All AR glasses and headsets work with some form of gaze.

Figure 17: Various AR headsets. [Top Left] Lenovo ThinkReality, [Top Right] Microsoft HoloLens, [Bottom Left] DreamWorld DreamGlass, and [Bottom Right] Meta's Meta 2. We use the HoloLens. [Image Sources: 117, 121, 35, 64]

Figure 18: Various AR glasses. These are less bulky in comparison to the headsets. However, they may also offer lesser capabilities. [Top Left] Epson Moverio BT-300, [Top Right] Magic Leap One, [Bottom] Google Glass Enterprise Edition. [Image Sources: 108, 36, 48].

There are multiple manufacturers of phones and tablets that offer touch screen based augmented reality. Apple's platform is ARKit, and is supported by the tablets and phones that have an A9 or A10 processor [7]. Google's software called ARCore is compatible with phones from a multitude of hardware companies including Apple, Oppo, Samsung and Sony among others [89]. Phone and tablet interactions are based on touching the screen (which is 2 dimensional), so interactions may *feel* limited to 2D with that hardware since dragging a finger up to move an object could be interpreted as moving up or moving back (in depth) [5]. However, being able to touch the interface provides the user with haptic feedback so that the user actually feels like they are touching something. This perhaps may be a valuable aspect for some users in terms of feedback [125]. Interactions on mobile and tablet AR apps are primarily based on touching the screen. However, embedding other voice control apps (such as SIRI) into the AR software development kits (such as ARKit) is foreseeable [88, 28].

Figure 19: Examples of touch screen based augmented reality. [Left] User interacts with the Houzz app on an iPhone via Apple's ARKit. [Right] User interacts with an educational app on an Acer tablet via Google's ARCore. [Image sources: 32, 1].

2.2 IoT Technology

As of 2019, there are an estimated 26.66 billion IoT connected devices worldwide with a steady increase projected for future years [115]. This thesis focusses on home based IoT devices. The devices range in capabilities. As well, devices may fall under the category of control points. Control points are IoT devices or applications through which users can control other IoT devices. Examples of control point IoT devices are the Google Home and Amazon Echo. Their Internet capabilities are used to manipulate or access other IoT devices such as smart televisions or thermostats. Control points like Google Home and Amazon Echo are based on voice input with commands like "Okay Google, Open What's new on Netflix" [87]. Phones and tablets also serve as control points via apps that are meant to control other devices. An example is the YouTube phone app that lets users cast videos on their smart TVs [42].

In addition, there are some platforms that let users program the way smart devices behave. A commercial example is the If This Then That (IFTTT) app that facilitates control of many smart devices and links them with services such as Instagram, Gmail, Spotify, etc. [53]. For example, the user may link up smart lights to turn on every time they receive an email. A research-based

example is the Reality Editor that is developed by *Huen et al.* in MIT's Fluid Interfaces Lab [50]. This is also one example that uses AR in conjunction with IoT. It is a phone-based AR app that lets the user “connect” IoT objects together to make them react to one another.

Figure 20: Example of a platform that lets users program the way IoT devices behave. Image of the Reality Editor AR app developed by *Huen et al.* from MIT Fluid Interfaces lab. Here, the user programs his car's thermostat to turn on when he leaves his office chair [50].

In terms of devices, the most common smart home ones are lights, thermostats, security cameras, speakers, televisions, washing machines, locks, vacuum cleaners, and printers/ scanners [29, 56]. Emerging smart devices include refrigerators, coffee tables, planters, plugs, small kitchen appliances (i.e. baking scales, cookers), microwaves, lawnmowers, sprinklers, beds, blinds, and fitness devices (i.e. BMI trackers) [16, 29, 56, 118]. There is an increase in the types of smart devices overall.

Figure 21: Examples of common IoT enabled home devices. [Top Left] Ecobee smart thermostat, [Top Center] iSmartAlarm security camera, [Top Right] Roomba smart vacuum, [Bottom Left] August smart lock, [Bottom Center] Bose smart speakers, [Bottom Right] Hotpoint smart washing machine. [Image sources: 29, 56, 31, 17]

Figure 22: Examples of emerging IoT enabled home devices. [Top Left] Click and Grow smart indoor garden, [Top Center] iHome smart plug, [Top Right] Anova Culinary Precision smart cooker, [Bottom Left] Velux smart blinds, [Bottom Center] Sobro smart coffee table, [Bottom Right] Husqvarna automower. [Image sources: 16, 29, 118, 112]

2.3 AR & IoT Combined Research

The research in the space of augmented reality in conjunction with the Internet of Things is based on a variety of different areas.

2.3.1 Education

IoT and AR are becoming increasingly ingrained into educational systems. Researchers are continuing to embed these concepts into education in various ways. XReality visualizes IoT networks using AR so that they are interactable [63]. This is used as an educational tool to inform students regarding sensor and computer networks. GAIA teaches children about energy efficiency through IoT and AR [81]. This project uses AR & IoT infrastructure to collect real-time energy data from a school building, and students use it to build projects via a lab kit. C-DAC researches and develops a range of AR and IoT products (AR Board, AR Book, AR Game) that are deployed into schools for practical use [11].

2.3.2 Home

Combined AR and IoT solutions for homes are another area of interest. ARIoT uses home scenarios to probe on the distribution of information to individual IoT objects (rather than having a central server) to access them on a need-to-know basis so as to increase the scalability of AR and IoT [58]. *Chang* develops interfaces for IoT objects with some home applications such as thermostats lights, wiring/ plumbing behind walls, fridge, and more [24]. ACADEMIC is a smart home control and energy monitoring interfaced with augmented reality that allows the user to view their real time energy consumption [97].

2.3.3 Smart Cities

The usage of AR and IoT for smart cities is another area of research interest. Some are concepts and others are potential implementations. *Chaves-Diéquez et al.* conceptualize a smart

city maintenance service to improve accessibility of sensor and actuator devices in the field in cases where responsiveness is vital [25]. *Pokrić et al.* create an AR application that is to be deployed in a smart city in Serbia in which there is an IoT infrastructure for data [95]. AR is triggered via image and geo-location markers and it is useful for aspects such as bus arrival times, bus routes, and tourist landmarks. In a different work, *Pokrić et al.* develop gaming technology to engage citizens regarding environmental issues using IoT and markerless AR methods [96]. *Rashid et al.* use AR and IoT to make smart cities more accessible for wheelchair users [101]. *Badouch et al.* research into the concept of IoT and AR for smart cities as a whole [13]. They note that some limitation factors do currently exist.

2.3.4 Environmental

AR and IoT are being used for engaging with the environment in terms of control and monitoring. *Chen et al.* probe China's coal production in terms of IoT smart mines and application of mobile devices [27]. MAYA uses AR and IoT to control the status of physical light [82]. Another study uses AR and IoT for smart maintenance and managements of road networks [21]. *Pokrić et al.*'s above discussed work is also one that focusses on air quality, temperature, air pressure and humidity monitoring [94]. They do this via an AR enabled IoT gaming concept.

2.3.5 User Interaction & Comfort

User interaction for AR and IoT is another space of focus. *Makolkina et al.* examine the relationship between a moving user and the density of IoT devices in the environment via an AR interface [72]. *Shin et al.* investigate allowing the user to create content in an AR and IoT interface [110]. *Son et al.* attempt to combat limitations of ARUI in terms of a lack of variability on meanings of things depending on contexts [114]. The proposal is an IoT platform side communication scheme, semantic communication as a core component, and its architecture based

on design considerations. *Alce et al.*'s study is based on seeing how users can discover, configure, and directly interact with multiple IoT devices [4]. Three AR interaction models are tested with two groups: low device density and high device density.

2.3.6 Backend Solutions

Some research work makes contribution towards the backend of AR and IoT. One AR middleware uses relational localization methods along with metadata information of sensors [14]. *Leppänen et al.* investigate an IoT system architecture in relation to web applications with mobile agents [65]. Another proposed architecture addresses the issue of perception of the smart environment [77]. *Rambach et al.* explore a 3D object tracking framework based on high quality 3D scans of objects [99].

2.3.7 Industrial / Market-Based

Specialized / industrial use cases in AR and IoT research have been investigated. This includes monitoring the stresses acting on loaded metal shelving [102]. *Rajan et al.* probe on machine fault diagnostics and condition monitoring with focus on temperature, pressure, speed, vibration noise, and more [100]. *Alam et al.* focus on improving the safety for maintenance tasks [3]. *Agrawal et al.* create an app for evaluation and visualization of a beating heart to make this doctor-patient interaction more intuitive [2]. A NASA-related work elaborates on best practices and guidelines for AR assisted procedure execution from user testing [60]. A research application for precision farming emphasizes on crop monitoring [93]. The shopping industry is another subject of research. Intelli-Mirror uses image processing techniques to detect a user and then displays a garment image on the person [68]. More specifically, it is for clothing and accessory display. Lastly, three key architectural components are identified to support scalable AR services

for IoT ready products [57]. This study finds that subjects clearly experience higher usability and greater satisfaction with AR-interactive shopping.

Figure 23: *Alce et al.* contribute three interaction models (floating icons, world in miniature, and floating menu). This is an image of their world in miniature model [4].

Figure 24: *Mylonas et. al* use AR and IoT for educational purposes. These are sample photos from the educational lab kit and AR tool in practice [81].

2.4 Elicitation Studies

Elicitation studies are conducted to gain insight into how the end-user wants to interact with the system. This insight allows innovators to create the interaction aspect of the product as per the user's wants. For example, multiple users may propose tapping on a touch screen slider rather than sliding it, and this may result in phone app developers to design sliders with taps. A vital portion of human computer interaction is determining what humans want in the technology they interact with and how they want to interact with it. Elicitation studies are a crucial part of this process.

2.4.1 *Gesture Elicitations*

Using gestures as input is by no means a new concept, and as far back as 1977, there have been attempts to track hand movements as input commands using a data glove [26]. Elicitations determine a list of referents for which the participant is asked to provide gestures. The list may include referents such as “insert”, “cut”, “accept” [122]. Usually, a final gesture set is determined by the amount of agreement amongst participants. This is further discussed in the agreement rate evaluations section.

2.4.2 *Voice Elicitations*

The use of natural interactions for technology is not a new concept by any means. However, there is a very limited amount of work that acknowledges voice command elicitations outside of multimodal elicitations. Most research that has to do with the elicitation of voice commands is a portion of multimodal elicitations. In total, there are 6 pieces of relevant research related to finding user preferences for voice interaction. *Hüttenrauch et al.* exclusively elicit voice commands for controlling mobile devices and services [52]. Their process includes an elicitation in which participants provide commands and a subsequent validation in which commands are given to

participants who convey their understanding of what the command would produce. Since the remainder of the elicitations were multimodal, the voice elicitation processes for those are covered in the following section.

Basic commands	
Confirm operation	Reject operation
Wake-up recognizer	Enter idle mode
Terminate service	Help
Transfer to human operator	Go to top level of service
List commands and/or functions	Cancel current operation
Go back to previous node/menu	Read prompt again
Digits	
Digits 0 to 9	Next digit repeated twice ("Double O")
Enter international access code	
Communications Commands	
Initiate digit dialling sequence	Dial a number or name
Home phone number (location)	Work phone number (location)
Mobile phone number (location)	Car phone number (location)
Personal number (attribute)	Make a call to the emergency services
Redial last dialled number	Set up a call-back to a called number
Accept incoming call	Reject incoming call
Forward incoming call	Set up a call diversion
Transfer an ongoing call	Put call on hold
Switch between two calls (hook flash)	Set up a conference call

Commands for the control of and navigation in media	
Play a recording	Start a recording
Stop temporarily	Resume interrupted playback
Stop playing a recording	Move forward faster than play
Move backward	Go to previous item
Go to next item	Provide more information about item
Modify item	Store item
Remove item	Respond to item
Forward item	Create new item
Send item	Move item to a new location
Reapply the undone action	Reverse the previous action
Commands for device and service settings	
List networks	Increase the volume
Decrease the volume	Silent mode
Re-activate the audio output	Silence the loudspeaker
Reactivate the loudspeaker	Silence the microphone
Re-activate the microphone	Activate vibrating alert
Deactivate vibrating alert	Summary of current device status
Change profile (pre-stored settings)	

Figure 25: *Hüttenrauch et al.*’s table depicts the commands covered by the ETSI standard (European Telecommunications Standards Institute) [52].

2.4.3 Multimodal Elicitations

Multimodal refers to the usage of multiple modes of user input. The focus for the purpose of this thesis is on the ones that involve gesture and voice/speech. Several of these elicitations also involved the **wizard-of-oz** technique [92, 83]. One work utilized a combination of methods to elicit gestures and voice commands: wizard-of-oz study, online survey, and video interview [92]. A few studies reported results via citing **number of occurrences** [83, 78]. One way of achieving this was counting number of occurrences for each interaction per referent i.e. 18 people did a two-hand pinch while 3 people said, “zoom in” to zoom in for *Nebeling et al.*’s study [83]. Most studies report a high-level view of the gestures that were involved with minimal details on which fingers

were used, etc. However, a 1993 study involved gestural details such as **finger usage** (i.e. how many and which) and **hand usage** (i.e. how many hands were used, whether there was a transition, etc.) [47]. Moreover, there were details on the voice commands portion such as **relevance of words** and **syntactic structure** [47]. *Peshkova et al.* [92] perform analysis via three **mental models** that participants use during control of an unmanned aerial vehicle (UAV):

- *imitative* (For gestures, parts of their body are directly mapped to movements of the UAV)
- *intelligent* (The operator expects some level of intelligence from the system. Applies to voice command and gestures. For gestures, a direction is indicated with a finger, etc.)
- *instrumented* (Gives the flight instructions using an imaginary tool. The tool may be an imaginary joystick, puppet ruler, or super power. Where super power is described as using arms or hands to push and pull as if via a super power).

One work contributed an elicitation tool called **Kinect Analysis** that intends to make multimodal elicitations with the Kinect hardware easier [84].

Multiple works aimed to find out which of the two methods were preferred by participants. One way of achieving this was **letting the user pick** which method (voice command or gesture) to use to perform an action [92, 47]. One more step to confirm preference was to **ask people** what they preferred [47]. **Consistency** was noted in that participants were using the same method (whichever they chose) repeatedly. One way of reaching a user-elicited set of gestures and commands was to: first gather the *total* proposed voice, gesture, and multimodal interactions; then determine the *distinct* proposed voice, gesture, and multimodal interactions; then determine the **maximum consensus** and **consensus distinct ratio** for gesture and voice by referent; then state the **number of occurrences** for each user-elicited interaction that reached the consensus threshold

of 3 [78]. Authors conclude that multimodal interactions did not reach a maximum consensus, and thus were not included.

Referent	Interaction	#
Open Browser	hand-as-mouse to select browser icon	8
	"open browser"	5
	"internet"	3
	"<browser name>" (e.g., "Internet Explorer," "Firefox," "Chrome")	3
Search Engine Query	"<query>"	6
	"search <query>"	5
Click Link	hand-as-mouse to select link	13
	"<link #>" (assumes all links have a number assigned to them)	3
	"back"	7
Go Back	flick hand from right to left	7
	hand-as-mouse to select back button	5
	flick hand from left to right	4
	"forward"	6
Go Forward	flick hand from right to left	5
	flick hand from left to right	5
	hand-as-mouse to select forward button	3
Open Link in Separate Tab	hand-as-mouse hovers on link until context menu appears, then hand-as-mouse to select menu option	3
Switch Tab	hand-as-mouse selects tab	7
	"next tab"	4
	"tab <#>" (assumes all tabs have a number assigned to them)	3
	flick hand	3
Find in Page	"find <query>"	4
	hand-as-mouse to select a find button, then type on virtual keyboard	3
Select Region	hand-as-mouse sweeps out diagonal of bounding box	6
	hand-as-mouse acts as highlighter, sweeping over each item to be included in region	3
	hand-as-mouse to select new tab button	6
Open New Tab	"new tab"	5
	"open new tab"	5
	"<url>" (e.g., "its2012conf.org")	7
Enter URL	type on virtual keyboard	5
	"go to <url>"	3
	"refresh"	9
Reload Page	"refresh page"	9
	move finger in spiral motion	3
Bookmark Page	hand-as-mouse selects bookmark button	7
	"bookmark page"	5
Close Tab	"close tab"	5
	hand-as-mouse to select close button on tab	4
	"close tab <#>" (assumes all tabs have a number assigned to them)	3
	hand-as-mouse to select close button on browser	6
Close Browser	"close browser"	3
	"exit"	3
	"exit all"	3

Figure 26: *Morris et al.*’s table depicts the number of occurrences for the voice and gesture interactions that reached the maximum threshold of 3 or above. This makes up the user-elicited interactions set for interacting with a web-browser in a living room.

One piece of work focused on a gestural elicitation but discussed the emergence of speech and multimodal interaction [61]. Some other studies reported participants unconsciously accompanying one method along with their main method of interaction [92, 47, 67]. In one work,

speech was noticed as an unconscious accompaniment to gestures [67]. In another, gestures were noticed as accompanying speech [47].

2.5 Agreement Rate Evaluations

Agreement is calculated by first determining if there is one gesture that is used by the most amount of people. Then, the number of users who agree with the most used gesture is calculated out of the total number of users. In one usage of an agreement score, if this number hits a certain threshold, that gesture can be concluded as natural amongst participants and can be added to the gesture set. There are several variations of this agreement rate calculation. A chunk of papers report using the **agreement rate** calculation derived by *Wobbrock et al.* [122].

$$A = \frac{\sum_{r \in R} \sum_{P_i \subseteq P_r} \left(\frac{|P_i|}{|P_r|} \right)^2}{|R|} \cdot 100\%$$

In this equation, r is a referent in the set of all referents R , P_r is the set of proposals for referent r , and P_i is a subset of identical symbols from P_r . The range of the equation is:

$1/|P_r| \cdot 100\% \leq A \leq 100\%$. The lower bound is non-zero because even when all proposals disagree, each one trivially agrees with itself. For example, if 15 out of 20 users proposed one gesture while 5 proposed another gesture for the same referent, r , $[(15/20)^2 + (5/20)^2] / 1 \cdot 100\% = 62.5\%$ is the agreement rate. Additionally, this agreement rate is higher than if than if 15/20 are of one form, 3/20 are of another, and 2/20 are of a third: $[(15/20)^2 + (3/20)^2 + (2/20)^2] / 1 \cdot 100\% = 59.5\%$. There is a reference to this in multiple gesture elicitation [107, 74, 20, 34, 39, 66]. In addition, some works also refer to the taxonomy developed by *Wobbrock et al.* [105, 85, 38].

However, there were a few gaps in this calculation with one being that there was a trivial agreement even when all proposals were unique, and thus *Vatavu et. al* revised this calculation [119].

$$\mathcal{AR}(r) = \frac{\sum_{P_i \subseteq P} \frac{1}{2} |P_i| (|P_i| - 1)}{\frac{1}{2} |P| (|P| - 1)}$$

The *AR* (Agreement Rate) notation is used to differentiate from the first calculation. For example, if it is assumed that there are 20 participants, from which $|P|=20$ proposals were collected for a given referent r , out of which 15/20 proposed one gesture and 5/20 propose another, i.e., $|P_1|=15$ and $|P_2|=5$. The number of pairs of participants in agreement with each other is $(15 \cdot 14)/2 + (5 \cdot 4)/2$, while the total number of pairs that could have been in agreement is $(20 \cdot 19)/2$. By dividing the two values, the agreement rate $AR(r) = (115/190) = .605$ is obtained. By comparison, the original calculation from *Wobbrock et al.* would yield $(15/20)^2 + (5/20)^2 = .625$. The amount of works that refer to this are also extensive [41, 22, 23, 109, 12, 66]. In addition, some of these works also refer to the elicitation method used by *Wobbrock et al.* [12]. However, these assume that each participant only proposes one gesture. Thus, to accommodate multiple proposals by each participant, there is also an extension of these calculations to generate a **maximum consensus** and that is also utilized extensively [33, 22, 23, 20]. The max consensus is defined by *Morris et al.* as “the percent of participants suggesting the most popular proposed interaction for a given referent” [78]. The formula can be viewed as:

$$\frac{\max_{P_i \subseteq P} |P_i|}{|P|}$$

where $|P|$ is the number of participants, P_i is a set of participants who made proposal i , so

$\max_{P_i \subseteq P} |P_i|$ is the number of participants who made the most popular proposal. However, the only

drawback in this approach lies in the case where there is no consensus among participants. For instance, if 2 people are asked for proposals and they each propose something different, the formula would yield $\frac{1}{2} = 50\%$ as the “max consensus” or “the percent of participants suggesting the most popular proposed interaction”. This is unexpected since there is, in fact, 0 agreement amongst the 2 participants.

Chen et al. points out a gap where there is no valid calculation for the case in which there are multiple proposals by the same participant as well as a 0 consensus amongst proposals of all participants [26]. The gap can be depicted by Chen’s table below:

<div>min value</div> <div><i>applicable study</i></div>	Never 0	0
<i>Single-proposal</i>	Wobbrock <i>et al.</i> ’s formula	Vatavu <i>et al.</i> ’s formula
<i>Multi-proposal</i>	Morris’ max-consensus	?

Figure 27: *Chen et al.*’s table depicts the gap for a valid calculation that accounts for multi-proposals by single participants in which the consensus amongst all participants’ proposal is 0. *Chen et al.* proposes a formula to fill this gap [26].

Chen et al. proposes a calculation to bridge this gap. *Chen* points this out as something that can be seen as a modification on *Morris et al.*’s maximum consensus.

$$AR = \frac{\max_{P_i \subseteq P} |P_i| - 1}{|P| - 1}$$

where $|P|$ is the number of participants, P_i is a set of participants who made proposal i , so

$\max_{P_i \subseteq P} |P_i|$ is the number of participants who made the most popular proposal. In this case, if 2 users proposed diverging inputs, the percentage would be $(1-1)/(2-1) = 0\%$. This thesis makes use of *Chen et al.*'s formula to ensure accounting for multiple proposals that may have no consensus. This formula is used for both gestures and voice commands in this thesis. The reason why *Chen et al.*'s method is used is because it is the most recently improved version of the agreement rate (as highlighted above).

Another way of depicting the most common voice commands and gestures (and multimodal input) is by **percentage** [92, 83]. Some of the works also branch off to *Morris et al.*'s [79] prescriptions such as the **production recommendation** [66]. Another point of common discussion is the **legacy bias** (which *Morris et al.* also focusses on) [79, 106, 22, 15]. Legacy bias refers to the predisposed ideas of participants- especially pertaining to their previous technological experiences- that may bias them when they are asked to propose a gesture. There are many open challenges in updating elicitation methods to incorporate production, priming, and partner techniques [79].

Some elicitations also rely on some level of **wizard-of-oz** techniques in which the user believes the system to be controlled by them while the researcher controls it [15, 67]. Some works specifically mention this in conjunction to removing the **gulf of execution** [67]. These techniques are employed to elicit what interactions users want without the limitations of the hardware/software.

2.6 Grounded Theory: Open Coding

Our work makes use of grounded theory: specifically, open coding in order to form qualitative conclusions. *Muller* [80] deeply discusses grounded theory. This allows the researcher to form conclusions regarding a domain without a dominant theory. It allows for a disciplined way to interpret and theorize. The order of data components occurs in the following order: taking data and open coding it to derive basic themes, going from open coding to axial coding to derive categories, going from axial coding to selective coding to derive concepts/dimensions, going from selective coding to the core concept to derive the emergent topic. Basically, it is a method to take raw data and gain meaning out of it via making note of emerging behaviours.

Open coding depicts the creation of a legend as data is being interpreted. For example, our study was analyzed by watching participation videos and encoding behaviours as they were observed i.e. if a new behaviour was noticed, it was added to the legend while if the same behaviour was repeated by participants, it was noted as being emergent. Examples of open codes in our study were the usage of each hand, the direction of movement, the direction of palm, the usage of each finger, the usage of words versus phrases versus sentences. The raw data as well as the open coding is available in the appendix. Interpretive consistency is maintained by having one person do all the open coding.

Chapter 3: User Elicitation Study for AR-IoT Controls

An elicitation study is conducted to determine user preferences for controlling IoT devices through headset AR. This chapter includes the study design, the classification of the elicited voice commands and gestures, and quantitative methods of analysis including the agreement rate. The study is an elicitation of gestural and voice command input based on 4 home based IoT scenarios. The goal of the study is to determine what kind of gestures and voice commands users want to use as well as which method of input they prefer.

3.1 Study Design

3.1.1 Pilot Study Participants

4 pilot study participants took part in the initial stages of the experiment (4 females). The pilot rounds were crucial in determining the final combination of tasks for the study. Most of these participants were from a Computer Science background while one was from a Communications and Media background. These participants were vital in pointing out a few changes that enhanced the quality of the experiment. Following the pilot study, details such as: the prompting script, timing and order of the tasks, as well as filming techniques were finalized.

3.1.2 Participants

16 volunteers participated in the main study (6 females, 9 males, 1 preferred not to say). Recruitment occurred through email lists, word of mouth, and snowball-sampling. The age range of participants is from 19 to 52 years (Mean = 27.125, Median = 23.5, SD = 8.085). Most of the participants came from a Computer Science background along with some from Math, Geomatics, and Mechanical Engineering. Of the 16 participants, 12 participants reported using remote controls to control a device at least once a week. 10 participants reported using voice commands to control a device at least once a week. 5 participants reported using AR *without* a headset while 4

participants reported using AR *with* a headset at least once a week. 5 people reported using VR *with* a headset at least once a week. In terms of immersive technology usage, 3 participants reported not using any at all while the rest gave examples such as Snapchat, gaming, development, etc. with platforms such as tablets, phones, VR and AR headsets. In terms of experience with IoT, 3 participants reported not using any at all while the rest provide examples such as development, sharing and capturing media, setting thermostat, displaying weather, controlling music, lights, and TVs on control platforms such as Chromecast, Zigbee, Google Home, Amazon Alexa, HoloLens, Kinect, phones. Although a couple of participants have developed using the HoloLens/Kinect, etc., no one has used headset AR to control devices in non-work-related setting. Detailed results of the pre-study questionnaire can be found in **Appendix C**.

Experience of Participants with Using Headset AR

Figure 28: A quarter of the participants use headset-based AR 2-5 times a week while the remaining 12 participants have not used it at all or used it very rarely. No one has used it to control IoT devices in a non-work-related setting.

3.1.3 Apparatus

The first-generation Microsoft HoloLens was used to display the *holographic* components of the experiment to the participants. *Holographic* refers to the virtual, graphical components that can only be seen through the headset. Although the study was conducted with the HoloLens

(limitations: field of view, weight, etc.), the results can likely be transferred to other AR headsets since the input is not dependent on the equipment. As previously motivated, removing the *gulf of execution* allows for any input to be acceptable [51]. No matter what gesture or voice command was implemented, the system would always work because this was a study in which the system was controlled by the researcher (albeit in front of the participant), thus it was kind of like a *wizard-of-oz* approach. This design provides the user the freedom to perform any interaction they want without any technology limitations of gesture recognition, command recognition, accent recognition, gaze-and-commit recognition, etc. Holographic remoting was implemented through a laptop so that the researcher could see what the participant saw and go to the next step when needed. Each front-end component on the holographic GUI was mapped to a keyboard key so that the researcher could press the corresponding key to trigger the next event. The participant's interaction was video and audio-captured using an iPhone camera mounted on a tripod. The camera captured both voice and gestural interactions.

Figure 29: Example of the *physical* and *holographic* components as seen by the wearer of the headset. In this IoT home control scenario, the video display is the physical component while the video 1 thumbnail is the holographic (or virtual, graphical) component. The physical components can be seen by everyone while the holographic components can only be seen by the wearer of the headset.

3.1.4 Scenarios

The goal is to determine how users want to control IoT devices in their homes using an AR headset. To accomplish this, home control scenarios had to be determined. First, a list of referents was created to determine the small actions (such as drag or click) that may be performed within bigger scenarios. Creating this list of actions aided in forming 4 scenarios: *Interacting with a Menu System*, *Environmental Control*, *Media Control*, and *Following a Workflow*. For each of these 4 scenarios, we created 2 different tasks. Tasks were created so as to incorporate various angles of usage which may alter the way controls are navigated. For example, do people prefer to do things differently when their hands are dirty? Storyboarding was done to create designs for each of these scenarios in which at least 3 alternative designs were considered for each (sample storyboard in **Appendix H**). For each task that the participant had to accomplish, they were told that they had to

complete the task using **voice commands** or **gesture** as decided by the research protocol. Furthermore, they were told that they had the option to complete the task **with or without** the use of the provided menu (not using the menu occurred very rarely). This attempts to cover situations in which participants would want to use gestures or voice commands that did not necessarily appear to be in sync with the specific design of the menu. For example, a user may either choose to repeatedly click on an “up” button to increase thermostat temperature or may instead want to raise their arm until a temperature is reached. Each of the 4 scenarios along with their respective 2 tasks are described in detail below.

3.1.4.1 Interacting with a Menu System

Interacting with a Menu System is selected as a scenario to determine how people want to interact when there are a series of steps that need to be completed to achieve a goal. This scenario is the result of the assumption that AR tools will use menu systems due to the observation that they are used by current technological control platforms all the time. A lot of current technology is based on the WIMP paradigm (windows, icons, menus, pointer) [62]. We expect IoT-based UIs to be similar. This scenario is specific to the cases where those steps are menu-like (something that would be seen on a computer or phone screen). These are designed as dynamic components that can be triggered by selecting a holographic component that sits next to a physical object in space.

For example, one task is connecting a computer to a print queue. Here the user sits in front of a monitor while wearing a HoloLens headset, and there is a holographic button next to the monitor. When this button is selected, the menu expands to allow the user to select whether the computer needs to be connected as an input or output device. Once *connect as input* device is selected, the menu expands with the next set of options and so on until the user is able to pick the appropriate printer. This dynamic design of the menu makes use of the virtually unlimited real-

estate that the HoloLens offers such that the menu gets bigger as more options appear. In this way, the user is always able to see all options and change their earlier selection if required. The interface sequence is observed in the figures below.

Figure 30: Connecting a Computer to a Print Queue Interface with the sequential steps. Initially, the holographic link button appears next to the physical computer.

Figure 31: Next, the user selects the link button: either via voice commands or gestures (as decided by the researcher).

Figure 32: All dynamic menus expanded. Once the link button is selected, the option to either *Connect as Input* or *Connect as Output* appears. Once the user selects *Connect as Input*, the options to select *Print Queue* or *Music* appears. Once *Print Queue* is selected, the list of available printers appears. The user selects *Printer 3*. The interface then indicates that the computer is “Connected to Printer 3”. Lastly, the user is asked to collapse the entire menu so that only the initial link button remains.

While preserving the interface design as much as possible, task 2 is created for this scenario: *Setting up a Lights Schedule*. This context requires the user to look up at a light that they would want to set up automation for. This automation consists of the days and times of when the light should be on. Once the user looks up at the light, they see a holographic clock “icon” next to the light. Once this icon is selected, the next options in the menu appear: *Time* and *Day*. One thing to note is that expanding a menu directly off the clock icon would cause the user to have to use the menu while craning their neck up, so the menu pops up directly in front of the user once the clock icon is selected. Once *Time* is selected, the user is able to use the dynamic menu to set up start and

end times. When *Day* is selected, the user is able to toggle which days they wanted the light to turn on. One difference between the two tasks is that the printer scenario is performed in an office chair in a sitting position while the lighting one is performed in a standing position. The amount of menu options/buttons is substantially more in this scenario than any other scenario. The interface sequence is observed in the figures below.

Figure 33: Setting Up a Lights Schedule Interface with the sequential steps. Initially, the holographic clock button appears next to the physical light. The user is looking up to see this.

Figure 34: Next, the user selects the clock button: either via voice commands or gestures (as decided by the researcher).

Figure 35: Once the clock is selected, the remainder of the menu will appear straight ahead so that the user does not have to keep craning their neck up. Once the options to pick either *Time* or *Day* appear, the user selects *Time* via voice or gesture.

Figure 36: Next, the options to select either *Start* or *End* time appear. The user selects *Start*. (Note that the small white circle follows the user's gaze, but that is not used as an input in our experiment.)

Figure 37: Next, the user makes use of the respective *Hours*, *Minutes*, *Seconds* sliders to set the start time. Then, the user toggles to either AM or PM.

Figure 38: Next, the user selects the *Done* button to indicate that they have finished entering the start time.

Figure 39: Next, the corresponding menu to enter *end* time appears (and the *start* time menu collapses). The sequence for entering *end* time occurs in the same way.

Figure 40: Next, the user selects *Day*. Then, the *time* menu collapses and the *day* menu will expand.

Figure 41: The user toggles on the days that they want the lights schedule to apply to (i.e. Monday). Next, they select *Done* to indicate that they have finished entering the days. Lastly, the user is asked to collapse the entire menu so that only the initial clock button remains.

3.1.4.2 Environmental Control

Environmental control is selected as a scenario to capture many current IoT home applications. Being able to control the physical environment is considered a major form of control. The menus are minimal in this scenario, so there are no dynamic menu components. One task is *blinds control* in which the task is to adjust blinds as prompted by the researcher. The user is provided with 3 buttons for this context (which they may or may not use to complete the task). They are first prompted to rotate the vertically-oriented blinds to the left. When completed, they are asked to rotate them to the right, and lastly, they are asked to open them entirely so that the window is fully uncovered.

Figure 42: Blinds control interface. A television display denotes a window where the paper with vertical lines is used to communicate the orientation of the blinds to the participants. The three buttons on the holographic component denote, respectively: rotate left, open/close window, rotate right.

Figure 43: The user holds down the *rotate right* button (and it turns black). The user was first asked to rotate blinds left, then rotate them to the right, then open them so that they are not covering the window any more.

The second task is thermostat control in which the user is first asked to lower the temperature to a certain degree, and then asked to increase the temperature to another number. Lastly, the user is asked to close the thermostat control. The interface styles between the 2 contexts are as consistent with each other as possible. The blinds control is performed in a seated position to stimulate sitting down on a couch while the thermostat control was performed standing up as if it can be done from any location while walking around in the house. One thing to note is that the blinds control is location-dependent while the thermostat control is not.

Figure 44: Thermostat Control Interface. Sequential steps as follows in the subsequent figures.

Figure 45: The user selects the *up* button to increase the temperature.

Figure 46: The user selects the *down* button to decrease the temperature.

Figure 47: The user selects the *close* button to close the thermostat control.

3.1.4.3 Media Control

Media control is selected as a scenario on the basis that a lot of situations that emerge from initial ideation are based on some sort of media transfer such as capturing/sharing/playing media. The focus of this media control is playing media. One task is *speaker control*. The location of these speakers is not specified: they could be imagined as embedded on the ceiling of the room so that they will play an audio piece when it is selected. In this context, the user is in a standing position, and the initial visual is a holographic audio thumbnail floating in the air in front of them. When the user selects this audio piece, the thumbnail is replaced by a progress bar that can be paused, played, or closed (to stop the audio). When the audio piece is fully “played”, the progress bar is again replaced by the thumbnail. The participant is first asked to select the audio piece to be played. Once it is playing, they are asked to pause it, and then play it again.

Figure 48: Speaker control interface. The speakers are at an unspecified location (or can be imagined as embedded on the ceiling). The user selects a holographic *audio thumbnail* to play it. Sequential steps as follows in the subsequent figures.

Figure 49: The *progress bar* replaces the audio thumbnail when that is selected. Here, the user selects the *pause button* to pause the song.

Figure 50: The *play button* is selected by the user to resume playing the song.

The other task is video display control in which the participant performs the task sitting down similar to the natural scenario of watching TV from a couch. The interfaces are almost identical in these two contexts. However, in the video context, the user is asked to select display 1 (out of 2 displays placed in different ends of the room) to eventually play the video. There is no interface component provided to the user to make this selection- they have to do it without a menu component. Once the correct display is selected, they are asked to start playing the video. Lastly, once the progress bar is moving along, they are asked to close the video player so that the video can stop playing. Amongst these two tasks, the speaker control is not location-dependent (the way that it is set up) while the video display control is.

Figure 51: Video control interface. The user makes it so that *Video 1* will play on *Display 1*. There are two displays identified to the user (one being on their left and another being on their right). They are asked to make it so that the video will play on Display 1 when they eventually decide to play it.

Figure 52: Next, the user picks the *video thumbnail* so that the video will actually play.

Figure 53: The *progress bar* replaces the video thumbnail when that is selected. The user closes the video player so that it will stop playing.

3.1.4.4 Following a Workflow

Following a workflow is selected as a scenario since accessing information in relevant contexts is also deemed a major advantage of AR. This scenario captures the idea that physical objects come with instructions to assemble or fix them, and that headset AR can be a steady alternative to paper or YouTube instructions. One task in this scenario is cooking a recipe in which participants are asked to assemble bread dough by following some instructions in AR. The interface is hovering above the “kitchen counter” in which there is a holographic image of the current ingredient. Along with this is an ingredient panel which displays the name of that ingredient as well as an instruction panel which indicates what to do in the current step (including required quantity). Right under these, there are two arrows which indicate/allow the user to navigate to the next step or go back to the previous step as required. There is no prompt given to the participant and they simply follow instructions until they reach the last step. This task is performed standing up (like when cooking at home) and most of the ingredients are located on the main “counter” while the flour is located on a counter behind the participant to mock a kitchen scenario.

Figure 54: Cooking bread dough interface with the sequential steps (in subsequent figures). This “kitchen counter” holds most ingredients that are needed. Initially, the user reads the first instruction, adds water, and selects the *next button* to go to the next step.

Figure 55: Next, the user reads the instruction, adds yeast, and selects the *next button* to go to the next step. (The *previous button* can be used to go to the previous step).

Figure 56: Next, the user reads the instruction, adds cooking oil, and selects the *next button* to go to the next step. In the same way, the user proceeds through the remaining steps until there are no more steps (*next button* is not available).

Figure 57: One of the steps require the user to go to a different counter to add flour (to simulate a kitchen scenario).

The second task is *fixing a boombox* and this scenario is performed sitting down as would likely be done in a real-life setting. The interface is identical with the only difference being that instead of ingredients, tools/steps to fix the boombox are indicated. As in the cooking task, participants are not prompted, but simply progress through the task at their own pace. One thing to note is that the participants' hands are occupied in both contexts, but they are actually immersed in dough in the cooking scenario (especially when the user starts kneading the dough).

Figure 58: Fixing a boombox interface. Subsequent figures show sequential steps. Initially, the user reads the instructions, turns the boombox backwards, and selects the *next button* to go to the next step.

Figure 59: The user reads the instructions and finds the correct screwdriver. From here, they can decide to use the *next button* to go to the next step or the *previous button* to go to the previous step.

Figure 60: The user reads the instruction, places the antenna on the boombox, and selects the *next button*. In the same way, the user progresses through the instructions until the last step (when the next button is no longer available).

Figure 61: The tools that the user has to use to fix the boombox sit on the counter.

3.1.5 Rounds

It is important to note that there are 16 combinations of tasks since all 8 tasks can be performed with either voice control or gesture control. Each participant was asked to take part in 8 combinations: where they performed each task only once with either voice or gesture control. This is done for the sake of time and for not making the participants perform the same tasks twice. The study is broken down into two rounds where round one consists of one task from each scenario: *connecting computer to print queue*, *blinds control*, *speaker control*, and *cooking*. Round two consists of the remaining 4 tasks (one from each scenario): *setting up a lights schedule*, *thermostat control*, *video display control*, and *fixing a boombox*. The order of completing the tasks is the same for everyone, and thus everyone experiences the same learning curve. Half of the participants perform round one with gesture navigation and round two with voice navigation. The other half of the participants perform round one with voice navigation and round two with gesture

navigation. This is done to avoid bias in terms of one navigation method over another. Also, this ensures that all tasks are tested for both methods. The breakdown of combinations and rounds can be observed in the figure below.

Combo #	Scenario	Task (Context)	Control Method
1	Interact with a Menu System	Computer to Print Queue	gesture
2	Interact with a Menu System	Computer to Print Queue	voice
3	Interact with a Menu System	Lights Schedule	gesture
4	Interact with a Menu System	Lights Schedule	voice
5	Environmental Control	Blinds	gesture
6	Environmental Control	Blinds	voice
7	Environmental Control	Thermostat	gesture
8	Environmental Control	Thermostat	voice
9	Media Control	Speaker Control	gesture
10	Media Control	Speaker Control	voice
11	Media Control	Video Display	gesture
12	Media Control	Video Display	voice
13	Follow a Workflow	Cooking	gesture
14	Follow a Workflow	Cooking	voice
15	Follow a Workflow	Fixing a Boombox	gesture
16	Follow a Workflow	Fixing a Boombox	voice

Figure 62: All possible 16 combinations of tasks. Half of the participants performed combinations 1-5-9-13 for round one and 4-8-12-16 for round two. The second half of the participants performed combinations 2-6-10-14 for round one and 3-7-11-15 for round two.

3.1.6 Procedure

Participants were asked to fill out a pre-study questionnaire in the beginning of the study (**Appendix A**). This pertains to their previous experience with controlling devices, AR, VR, and IoT. Next, they were read out initial instructions and fitted with the headset to make sure that things were visible, and it was comfortable. They were then read out round one instruction in which they were told which control method to use. Any questions they had were answered. After completing

the 4 tasks in this order: *computer to print queue*, *blinds*, *speakers*, and *cooking*, they washed their hands that were covered in bread dough. While this occurred, things were set up for round two. After the participant came back, the round one semi-structured interview took place. The interview was recorded as well. It consisted of questions pertaining to the system as well as the control method that they had been asked to use for round one. After this, the participant was fitted with the headset again, and read the round two instructions. This time they were asked to use the other method of control (gesture if they had used voice in round one and vice versa). In round two, they completed the remaining 4 tasks in this order: *lights schedule*, *thermostat*, *video display*, *fixing the boombox*. After the end of round two, the same semi-structured interview occurred for the second method of control. At the end of those questions, another semi-structured interview occurred with the purpose of comparing and contrasting the two methods of control. The researcher's script along with the semi-structured interview can be found in **Appendix B**. At the end, the participant was thanked with some chocolate as a token of appreciation. After the participant left, notes were written down and participation videos were uploaded to a computer.

3.2 Data Analysis

The 16 participants were prompted for input 49 times resulting in a total of 784 inputs. Out of which there were 392 gestures and 392 voice commands input by the users. **Appendix D** contains the full data set of elicited gestures and voice commands. These were classified, and the methodology is explained in detail in the following section. From the resulting set of gestures and voice commands, agreement rates between participants were calculated and interpreted. A consensus set, as defined by *Chen et al.*'s agreement rate [26], is presented for headset augmented reality gestures and voice commands. The result of the method of preference is also discussed.

3.2.1 Observational Methodology & Classification of Input

Muller's [80] work demonstrates that grounded theory is a core method of qualitative research in which open coding is a central concept. Open coding pertains to attaching “codes” or concepts to the observed data. This approach is used for the purpose of this thesis. Initially, each of the 784 inputs were recorded on video and a short description was written for each one of them in a table. For gestures, this included observation on:

- which hand was used
- which fingers were used
- which way the palm was facing
- what kind of action it was
- Additional things were also noted such as:
 - when a participant decided to supplement their gesture with a voice command
 - whether a hand was occupied if there was a switch in hands
 - whether the entire arm was used
 - any additional context

For voice, the short description was a quotation of the words they used. Other contextual and supplemental observations were also made.

Next, codes were attached to behaviour. Codes were developed by identifying key recurring behaviours that were observed. A legend was created for these codes. The derived gesture set, voice command set, and method preference is discussed in this chapter. The remaining observations are a part of chapter 4.

3.2.1.1 Gesture Set

12 unique gestures from 392 gesture inputs were initially identified. These would later be narrowed down to form a gesture set.

Figure 63: 12 unique gestures identified out of 392 gesture inputs. These gestures were sometimes performed with the entire arm stretched or with both hands, and these details are indicated in the raw data in Appendix D. A combination of gestures was also used sometimes. A gesture set was narrowed down from these gestures.

To determine a gesture set, *Chen et al.*'s method was used in which multiple proposals from a single participant and the possibility of a 0 consensus are both accounted for. Sometimes a referent is used in both tasks; in other cases, it is used in only one of the two tasks. When a referent is used in one task, it means 8 participants made proposals for it, but when a referent is used in both tasks, it means 16 participants made proposals for it. The agreement rate is determined either out of 8 or 16 accordingly.

In some cases, participants proposed a combination of gestures as one gesture. In this case, each individual gesture was regarded *atomically* where the combination of atomic parts created one proposal. For example, when asked to play a holographic video thumbnail on a physical display, one participant chose to: **tap** on the hologram, then **drag** it toward the display, then

release it in that direction. Thus, the three atomic parts of this gesture were: tap, drag, and release. In cases where combination gestures were present, the combinations were often varying amongst participants. However, common atomic parts still existed. Thus, results were gathered with a focus on atomic gestures. For instance, it was noted that in this example, drag was the common factor that was used most often amongst participants. Looking at this from a conservative approach, it can be inferred that at the least, it is natural for most participants to drag the holographic video to the display. Thus, drag can be considered a main gesture for this interaction (that may be supplemented with other gestures).

Participant	Display 1
2	swipe->L A-L [R] ARM
4	tap and drag->L and release T, P [R]
6	tap and drag->L, P [R]
8	tap and drag->L and release P [R]
10	swipe->L A-L [R] ARM
12	swipe->L A-L [R] ARM
14	air hold and drag->L and release A-F, T, F4 [R]
16	air hold A-F, T, F4, and drag->L P [R]

Figure 64: Atomic gestures often made up one gestural proposal. It was noted that drag was the most common factor amongst participants to perform this task.

On multiple occasions, there were cases where a referent was used multiple times within the same task. For instance, the participant was required to ***pick a button*** three times when connecting a computer to a print queue (i.e. there were three buttons). This meant that the participant contributed three proposals for the same referent. The participant could perform the same or different gestures in each of the three proposals (they were not instructed regarding this in any way). In mathematical terms, the three proposals would be called a ***bag***. A bag is a group of values where there may or may not be repetition.

Referent	Generic Participant Number		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
Pick Button (Interacting with Menu System)	Pick Printer Menu Buttons	Proposal 1	air tap	tap	air tap	air tap	tap	tap	swipe	tap	Proposed Voice Commands							
		Proposal 2	air tap	tap	air tap	air tap	tap	tap	tap	tap								
		Proposal 3	air tap	tap	air tap	air tap	tap	tap	tap	tap								
	Pick Light Menu Buttons	Proposal 1	Proposed Voice Commands								tap		air tap	tap	tap	air tap	tap	tap
		Proposal 2									tap		air tap	tap	tap	air tap	tap	tap
		Proposal 3									swipe		air tap	tap	tap	air tap	tap	tap
		Proposal 4									swipe		air tap	tap	tap	air tap	tap	tap
		Proposal 5									tap		air tap	tap	tap	air tap	air hold & tap	tap
		Proposal 6									air hold & tap		air tap	tap	tap	air tap	air hold & tap	tap

Figure 65: Each participant’s bag of proposals. All proposals for each participant make up the bag for that participant. For example, participant 1’s bag is indicated with yellow highlighter: “air tap, air tap, air tap”. Participant 7’s bag is indicated with pink: “swipe, tap, tap”.

To further analyze this, a *set* was determined. A set is a group of values where there may not be repetition. To achieve this, each participant’s bag was reduced to a set.

Referent	Generic Participant Number		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
Pick Button (Interacting with Menu System)	Pick Printer Menu Buttons	Proposal 1	air tap	tap	air tap	air tap	tap	tap	swipe	tap	Proposed Voice Commands							
		Proposal 2	air tap	tap	air tap	air tap	tap	tap	tap	tap								
		Proposal 3	air tap	tap	air tap	air tap	tap	tap	tap	tap								
	Pick Light Menu Buttons	Proposal 1	Proposed Voice Commands								tap		air tap	tap	tap	air tap	tap	tap
		Proposal 2									tap		air tap	tap	tap	air tap	tap	tap
		Proposal 3									swipe		air tap	tap	tap	air tap	tap	tap
		Proposal 4									swipe		air tap	tap	tap	air tap	tap	tap
		Proposal 5									tap		air tap	tap	tap	air tap	air hold & tap	tap
		Proposal 6									air hold & tap		air tap	tap	tap	air tap	air hold & tap	tap
	Participant's Set		air tap	tap	air tap	air tap	tap	tap	swipe, tap	tap	tap, swipe, air hold		air tap	tap	tap	air tap	tap, air hold	tap

Figure 66: Each participant’s bag of proposals is narrowed down to a set.

Then, a final set was determined by combining all the participants’ set. This is the set of all gestures used by any participant.

Referent	Generic Participant Number		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
Pick Button (Interacting with Menu System)	Pick Printer Menu Buttons	Proposal 1	air tap	tap	air tap	air tap	tap	tap	swipe	tap	Proposed Voice Commands							
		Proposal 2	air tap	tap	air tap	air tap	tap	tap	tap	tap								
		Proposal 3	air tap	tap	air tap	air tap	tap	tap	tap	tap								
	Pick Light Menu Buttons	Proposal 1	Proposed Voice Commands								tap		air tap	tap	tap	air tap	tap	tap
		Proposal 2									tap		air tap	tap	tap	air tap	tap	tap
		Proposal 3									swipe		air tap	tap	tap	air tap	tap	tap
		Proposal 4									swipe		air tap	tap	tap	air tap	tap	tap
		Proposal 5									tap		air tap	tap	tap	air tap	air hold & tap	tap
		Proposal 6									air hold & tap		air tap	tap	tap	air tap	air hold & tap	tap
	Participant's Set		air tap	tap	air tap	air tap	tap	tap	swipe, tap	tap	tap, swipe, air hold		air tap	tap	tap	air tap	tap, air hold	tap
	Final Set		air tap, tap, swipe, air hold															

Figure 67: The final set is determined by combining all the participants’ sets.

Once the final set was determined, each gesture was numbered by how many participants had that gesture in their own set (how many participants ever proposed the gesture).

Referent	Generic Participant Number								P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
Pick Button (Interacting with Menu System)	Pick Printer Menu Buttons	Proposal 1	air tap	tap	air tap	air tap	tap	tap	swipe	tap	Proposed Voice Commands													
		Proposal 2	air tap	tap	air tap	air tap	tap	tap	tap	tap														
		Proposal 3	air tap	tap	air tap	air tap	tap	tap	tap	tap														
	Pick Light Menu Buttons	Proposal 1	Proposed Voice Commands														tap	air tap	tap	tap	air tap	tap	tap	tap
		Proposal 2															tap	air tap	tap	tap	air tap	tap	tap	tap
		Proposal 3															swipe	air tap	tap	tap	air tap	tap	tap	tap
		Proposal 4															swipe	air tap	tap	tap	air tap	tap	tap	tap
		Proposal 5															tap	air tap	tap	tap	air tap	air hold & tap	tap	tap
		Proposal 6															air hold & tap	air tap	tap	tap	air tap	air hold & tap	tap	tap
	Participant's Set		air tap	tap	air tap	air tap	tap	tap	swipe, tap	tap	tap, swipe, air hold													
	Final Set		air tap, tap, swipe, air hold																					
	Final Set By Number of People who ever proposed gesture		tap (11), air tap (5), swipe (2), air hold (2)																					

Figure 68: The final set is numbered by the amount of people who ever proposed the gesture.

Lastly, the gesture proposed by most amount of people was declared, and its agreement rate was calculated.

Referent	Generic Participant Number		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16			
Pick Button (Interacting with Menu System)	Pick Printer Menu Buttons	Proposal 1	air tap	tap	air tap	air tap	tap	tap	swipe	tap	Proposed Voice Commands										
		Proposal 2	air tap	tap	air tap	air tap	tap	tap	tap	tap											tap
		Proposal 3	air tap	tap	air tap	air tap	tap	tap	tap	tap											tap
	Pick Light Menu Buttons	Proposal 1	Proposed Voice Commands									tap	air tap		tap	tap	air tap	tap	tap		tap
		Proposal 2										tap	air tap		tap	tap	air tap	tap	tap	tap	tap
		Proposal 3										swipe	air tap		tap	tap	air tap	tap	tap	tap	tap
		Proposal 4										swipe	air tap		tap	tap	air tap	tap	tap	tap	tap
		Proposal 5										tap	air tap		tap	tap	air tap	air hold & tap	tap	tap	tap
		Proposal 6										air hold & tap	air tap		tap	tap	air tap	air hold & tap	tap	tap	tap
	Participant's Set		air tap	tap	air tap	air tap	tap	tap	swipe, tap	tap	tap, swipe, air hold	air tap	tap	tap	air tap	tap, air hold	tap	tap			
	Final Set		air tap, tap, swipe, air hold																		
	Final Set By Number of People who ever proposed gesture		tap (11), air tap (5), swipe (2), air hold (2)																		
	Gesture Proposed by most amount of People		tap (0.67 agreement rate)																		

Figure 69: The gesture proposed by the most amount of people is declared and is supplemented with its agreement rate.

The remaining gesture analysis charts are in **Appendix E**.

Referent	Gesture	Agreement Rate
<i>Interacting with Menu System</i>		
Expand Menu	tap	0.53
Pick Button	tap	0.67
Set Slider	drag	1
Set Toggle	tap	0.71
Collapse Menu	tap	0.29 *low agreement rate
<i>Environmental Control</i>		
Pick Directional Button/ Specify Direction	tap	0.53
Open Blinds Entirely	swipe	0.71
Close the Control Panel	tap	0.86
<i>Media Control</i>		
Select Media to Play	tap	0.6
Select Button/Modify Playing Status	tap	0.67
Select Physical Display for Media	drag	0.57
<i>Following a Workflow</i>		
Go to Next/ Previous Step (not enough users used previous)	tap	0.6

Figure 70: Overview of final gesture set. It should be noted that the “collapse menu” referent had a very low agreement rate for gestural input.

Based on observations, a system design could use the following 3 gestures to live up to user preferences: tap, drag, and swipe. Tap is the most common and is seen across all 4 scenarios. Users opted for this when there was some graphical component on the field of view that they could “select”. Examples of this are generic buttons (such as those users see on their smartphones and computers), graphical icons (that are meant to expand into menus), toggles (that appear as checkboxes or dichotomous buttons), or directional arrows/indicators (that sit on top of buttons). Drag was used when there was some physical movement implied: one being the case of the graphical slider (such as those users may use to turn up the volume or brightness on their phones), and the other being “moving” a graphical media thumbnail to a physical TV/Computer to play it. Swipe was used when there was some physical movement that was like the movement of an actual physical object: opening “blinds” like one would “draw curtains” with both arms. Swiping is a

gesture commonly used on smartphones to control features such as “close tabs”. The difference found in the swipe used to control the curtains was that the entire hands and arms were used in order to accomplish the task. This could be an example of how people’s knowledge of how they interact with technology may be merging with how they deal with everyday tangible objects. This is further discussed in Chapter 4.

3.2.1.2 Voice Commands Set

We contribute a novel method for analyzing voice commands called ***Voice Command Pattern Template***. This is a major contribution of this thesis since this kind of analysis has not been done before. Voice commands are very flexible. We want to determine if there is a pattern that people use when they try to use a voice command for IoT interactions. We are inspired by grammars and attempt to develop a grammar for voice commands. Grammars that we want are simple and IGNORE sequence. Some components can be skipped. We heuristically state that a user is following the grammar when they use more than 50% of the components. The voice commands set was found by looking for commonalities in the voice commands that each participant used to accomplish the same result. This was done for each time a voice command was required.

Our analysis technique results in a pattern for the command. An example of this is the **Pick Directional Button/ Specify Direction** referent for the Environmental Control scenario. Raw data from 8 participants that were tasked with changing the orientation of the blinds in the left direction:

P1	"move to left"
P2	"rotate to the left using the button"
P3	"rotate blinds to the left"
P4	"rotate the blinds to the left"
P5	"click"
P6	"spin left"
P7	"blinds rotate to the left"
P8	"rotate blinds to left"

Initially, the most common factor that was seen is the usage of “[to the] left”.

P1	"move to left"
P2	"rotate to the left using the button"
P3	"rotate blinds to the left"
P4	"rotate the blinds to the left"
P5	"click"
P6	"spin left"
P7	"blinds rotate to the left"
P8	"rotate blinds to left"

Next, it was noted that the word “blinds” is used quite often.

P1	"move to left"
P2	"rotate to the left using the button"
P3	"rotate blinds to the left"
P4	"rotate the blinds to the left"
P5	"click"
P6	"spin left"
P7	"blinds rotate to the left"
P8	"rotate blinds to left"

Lastly, it was noted that *verbs* like “rotate” and “spin” or “move” were used various times.

P1	"move to left"
P2	"rotate to the left using the button"
P3	"rotate blinds to the left"
P4	"rotate the blinds to the left"
P5	"click"
P6	"spin left"
P7	"blinds rotate to the left"
P8	"rotate blinds to left"

Since 7 people used at least 2 of 3 criteria ($\geq \frac{1}{2}$ of the criteria) as highlighted in the 3 colours, they were counted as using the same general formula of:

“rotate”/other verb (“spin”/ “move”) + “blinds” + “[to the] left” (7 of 8 users)

where “rotate” was the most used verb and “spin” and “move” were synonymous. It should be noted that sequence of components does not matter i.e.) P7 said “blinds” first. Next, the voice commands from proposal 2 (orienting blinds to the right direction) were also analyzed in the same way. The formula derived from that was the following (where the same 7 people used the formula):

“rotate”/other verb (“spin”/ “move”) + “blinds” + “[to the] right” (7 of 8 users)

It was noted that “to the” are filler words only used by some users, and thus were dropped. From here, looking at both proposals (and removing any uncommon factors), a grammar for the user-preferred command was derived. It was inferred that if a system accepts a command that takes in:

“rotate”/other verb (“spin”/ “move”) + “blinds” + “left”/ “right”

(in non-specific order) there is a high chance that it will coincide with what words the users would intuitively use. The total number of *unique individuals* who used this formula was **7 out of 8**.

Before a conclusion was derived, the same referent (**Pick Directional Button/ Specify Direction**) from the corresponding Environmental Control scenario task (thermostat control) was evaluated. The remaining set of 8 people were tasked with changing the room temperature to 24 degrees. This was like the blinds since both required the user to either use directional buttons (left/right for blinds and up/down for temperature) or interact in any other way to accomplish the task (such as imagining a rotating dial). The raw data for the temperature task was:

P9	"Hey Tim I'm cold can we change the hold at temperature to 24?"
P10	"select hold at 24"
P11	"hold at 24 degrees Celsius"
P12	repeatedly saying "plus"
P13	"hey Siri change the hold at temperature to 24"
P14	"change hold at 19 to 24"
P15	"set the temperature to 24"
P16	"set up hold at 24 Celsius"

Initially, the most common factor that was seen is the usage of “hold at [temperature]”.

P9	"Hey Tim I'm cold can we change the hold at temperature to 24?"
P10	"select hold at 24"
P11	"hold at 24 degrees Celsius"
P12	repeatedly saying "plus"
P13	"hey Siri change the hold at temperature to 24"
P14	"change hold at 19 to 24"
P15	"set the temperature to 24"
P16	"set up hold at 24 Celsius"

Next, it was noted that “[to] 24” is used quite often.

P9	"Hey Tim I'm cold can we change the hold at temperature to 24?"
P10	"select hold at 24"
P11	"hold at 24 degrees Celsius"
P12	repeatedly saying "plus"
P13	"hey Siri change the hold at temperature to 24"
P14	"change hold at 19 to 24"
P15	"set the temperature to 24"
P16	"set up hold at 24 Celsius"

Lastly, it was noted that *verbs* like “change” and “set” or “select” were used various times.

P9	"Hey Tim I'm cold can we change the hold at temperature to 24?"
P10	"select hold at 24"
P11	"hold at 24 degrees Celsius"
P12	repeatedly saying "plus"
P13	"hey Siri change the hold at temperature to 24"
P14	"change hold at 19 to 24"
P15	"set the temperature to 24"
P16	"set up hold at 24 Celsius"

Since 7 people used at least 2 of 3 criteria ($\geq \frac{1}{2}$ of the criteria), they were counted as using the same general formula of:

“change”/ other verb (“set”/ “select”) + words on UI for temperature value (“hold at”)
+ “[to] 24” (7 of 8 users)

where “change” is the most used verb and “set” and “select” are synonymous verbs. Next, the voice commands from proposal 2 (changing temperature to 18) were also analyzed in the same way. The formula derived from that was the following (where the same 7 people used the formula):

“change”/ other verb (“set”/ “select”) + words on UI for temperature value (“hold at”)
+ “[to] 18” (7 of 8 users)

From here, looking at both proposals (and removing any uncommon factors), a grammar for the user-preferred command was derived. It was inferred that if a system accepts a command that takes in:

“change”/ other verb (“set”/ “select”) + words on UI for temperature value (“hold at”)
+ numeric value

(in non-specific order) there is a high chance that it will coincide with what words the users would intuitively use. The total number of *unique individuals* who used this formula was **7 out of 8**.

After this, the above two formulas for the referent (Pick Directional Button/ Specify Direction) were examined:

“rotate”/other verb (“spin”/ “move”) + “blinds” + “left”/ “right”

“change”/ other verb (“set”/ “select”) + words on UI for temperature value (“hold at”)
+ numeric value

After removing uncommon factors, it was determined that a super-formula that encompasses both would be:

**"rotate"/"change"/ other verb ("spin"/"move"/"set"/"select")
+ object ("blinds"/ UI words for temperature)
+ value ("left"/"right"/numeric value)**

Next, *Chen et al.*'s agreement rate was calculated in the following way:

7 individuals from Task 1 + 7 individuals from Task 2 = 14 individuals who used the formula

Where there were 16 individuals. Thus: $\frac{14-1}{16-1} = 0.87$

In this way, voice command formulas and agreement rates were determined for all 4 scenarios.

An example of the voice command analysis is below. The remainder voice command analysis tables can be found in **Appendix F**.

Referent	Generic Participant Number	P9	P10	P11	P12	P13	P14	P15	P16	Emerging Command	Number of Users
Pick Button (Interacting with Menu System)	Pick Printer	Proposal 1	"connect as input"	"connect as input"	"select connect as input"	"connect as input"	"click"	"connect as input"	"connect as input"	"Connect as input"	7 out of 8
	Menu Buttons	Proposal 2	"connect as input print queue"	"print queue"	"select print queue"	"print queue"	"click"	"print queue"	"print queue"	"Print Queue"	7 out of 8
		Proposal 3	"printer 3"	"printer 3"	"select printer 3"	"printer 3"	"select printer 3"	"printer 3"	"printer 3"	"Printer 3"	8 out of 8
	Generic Participant Number	P1	P2	P3	P4	P5	P6	P7	P8		
	Proposal 1	"hey Tim set time"	"time"	"time"	"time"	"time"	"time"	"set the alarm (time)"	"set time"	"Time"	7 out of 8
	Proposal 2	"start time tim"	"start"	"start"	"start"	"start"	"start"	"start"	"start"	"Start"	8 out of 8
	Proposal 3	"done"	"done"	"done"	"done"	"that's it"	"done"	"turn off"	"done"	"Done"	6 out of 8
	Proposal 4	"end"	"end"	"end"	"end"	"end time"	"end"	"turn off at: "	"set up end time"	"End"	7 out of 8
	Proposal 5	"day tim"	"done"	"done" day"	"done" day"	"and we're done" day"	"done" day"	"turn it off"	"done" "set up day"	"Day"	6 out of 8
	Proposal 6	"done"	"done"	"done"	"done"	"that's it"	"done"	"go to time"	"done"	"Done"	6 out of 8
	Final Command - based on all proposals (& Number of Users Who Ever Proposed It)									Exact words on Button	16 out of 16
	Agreement Rate									1	

Figure 71: Voice command analysis table for picking a button referent. The final command and agreement rate emerge by separately analyzing both tasks and combining their results.

Referent	Generic Participant Number	P9	P10	P11	P12	P13	P14	P15	P16	Emerging Command	Number of Users		
Go to Next / Previous Step (Following a Workflow) Note: Not enough usage of previous	Cooking a Recipe	Proposal 1	"ok next"	"next"	"ok next step"	"ok next"	"k next step"	"next step"	"next"	"uh next"	"[ok] + 'next' + 'step'"	5 out of 8	
		Proposal 2	"next"	"next"	"next step"	"next"	"next step"	"next step"	"next"	"uh next"	"'next' + 'step'"	8 out of 8	
		Proposal 3	"ok next"	"next"	"ok next step"	"ok next"	"next step"	"next step"	"next"	"uh next"	"[ok] + 'next' + 'step'"	5 out of 8	
		Proposal 4	"ok next"	"next"	"k next step"	"next"	"next step"	"next step"	"next"	"uh next"	"'next' + 'step'"	8 out of 8	
		Proposal 5	"k next"	"k next"	"k next step"	"next"	"next step"	"next step"	"next"	"uh next"	"[ok] + 'next' + 'step'"	5 out of 8	
		Proposal 6	"next"	"k next"	"next step"	"ok next"	"next step"	"last step go back" "next step"	"next step"	"next"	"ok next"	"[ok] + 'next' + 'step'"	6 out of 8
		Proposal 7	"next"	"next"	"yeah next step"	"next"	"next step"	"next step"	"next"	"ok next"	"'next' + 'step'"	8 out of 8	
	Task's Emerging Command										"'next' + 'step'"		
	Fixing a Boombox	Proposal 1	"k tm go to the next step"	"uh next"	"k next step"	"next"	"hey google I wanna fix this boombox"	"next"	"go to the next one"	"next"	"'next' + 'step'"	8 out of 8	
		Proposal 2	"next step tm"	"next"	"next step"	"next"	"ok next step whats next"	"next"	"go to the next one"	"next"	"'next' + 'step'"	8 out of 8	
Proposal 3		"next step tm"	"next"	"next step"	"next"	"whats the next step"	"next"	"go to the next one"	"k next"	"'next' + 'step'"	8 out of 8		
Proposal 4		"next step tm"	"next"	"next step"	"next"	"whats next"	"next"	"go to the next one"	"done next"	"'next'"	8 out of 8		
Proposal 5		"tm can you go back"	"next"	"next step"	"next"	"whats next"	"next"	"next"	"next"	"'next'"	8 out of 8		
Proposal 6		"next step tm"	"next"	"next step"	"next"	"ok whats next"	"next"	"next"	"ok next"	"'next'"	8 out of 8		
Proposal 7		"next step tm"	"next"	"next step"	"next"	"done that whats next"	"back" "next"	"next"	"yeah next"	"'next'"	8 out of 8		
Proposal 8		"next"	"next"	"next step"	"next"	"whats next"	"next"	"next"	"yeah done next"	"'next'"	8 out of 8		
Task's Emerging Command										"'next'"			
Final Command - based on all proposals (& Number of Users Who Ever Proposed It)										"'next'"	16 out of 16		
Agreement Rate										1			

NOTE: "ok" was not meant as a part of the command, but since it emerged amongst multiple participants, it was analyzed as well. This is further discussed in chapter 4

Figure 72: Voice command analysis table for going to next/previous step referent. This is a more complex example of analysis.

A few rules that were used to develop this technique include:

- 1) Arbitrarily, more than 25% of users have to propose something in the same proposal round for it to be considered (at least 3 out of 8 in this case).

Proposal 6	"next"	"k next"	"next step"	"ok next"	"next step"	"last step go back" "next step"	"next step"	"next"	"ok next"	"[ok] + 'next' + 'step'"	5 out of 8
Proposal 7	"next"	"next"	"yeah next step"	"next"	"next step"	"next step"	"next"	"next"	"ok next"	"'next' + 'step'"	8 out of 8

Figure 73: "Ok" is used at least 3 times in proposal 6, so it appears as a part of the emerging command. It is only used once in proposal 7, so it does not appear there.

- 2) If there are uncommon factors amongst the emerging command proposals of one task, those factors are removed from the emerging command of that task.

Referent	Generic Participant Number	P9	P10	P11	P12	P13	P14	P15	P16	Emerging Command	Number of Users
Cooking a Recipe	Proposal 1	"ok next"	"next"	"ok next step"	"ok next"	"k next step"	"next step"	"next"	"uh next"	"[ok] + 'next' + 'step'"	5 out of 8
	Proposal 2	"next"	"next"	"next step"	"next"	"next step"	"next step"	"next"	"uh next"	"'next' + 'step'"	8 out of 8
	Proposal 3	"ok next"	"next"	"ok next step"	"ok next"	"next step"	"next step"	"next"	"uh next"	"[ok] + 'next' + 'step'"	5 out of 8
	Proposal 4	"ok next"	"next"	"k next step"	"next"	"next step"	"next step"	"next"	"uh next"	"'next' + 'step'"	8 out of 8
	Proposal 5	"k next"	"k next"	"k next step"	"next"	"next step"	"next step"	"next"	"uh next"	"[ok] + 'next' + 'step'"	5 out of 8
	Proposal 6	"next"	"k next"	"next step"	"ok next"	"next step"	"last step go back" "next step"	"next step"	"ok next"	"[ok] + 'next' + 'step'"	5 out of 8
	Proposal 7	"next"	"next"	"yeah next step"	"next"	"next step"	"next step"	"next"	"ok next"	"'next' + 'step'"	8 out of 8
Task's Emerging Command										"'next' + 'step'"	

Figure 74: "Ok" is not used as a part of emerging command for all proposals. However, "next" + "step" is a common factor throughout. Thus, "ok" is removed from the task's emerging command.

- 3) Furthermore, if there are uncommon factors amongst the emerging command of the two tasks, those factors are removed from the final command.

Referent	Generic Participant Number	P9	P10	P11	P12	P13	P14	P15	P16	Emerging Command	Number of Users
Go to Next / Previous Step (Following a Workflow) <i>Note: Not enough usage of previous</i>	Cooking a Recipe	Proposal 1	"ok next"	"next"	"ok next step"	"ok next"	"k next step"	"next step"	"next"	"uh next"	[ok] + "next" + "step" 5 out of 8
		Proposal 2	"next"	"next"	"next step"	"next"	"next step"	"next step"	"next"	"uh next"	"next" + "step" 8 out of 8
		Proposal 3	"ok next"	"next"	"ok next step"	"ok next"	"next step"	"next step"	"next"	"uh next"	[ok] + "next" + "step" 5 out of 8
		Proposal 4	"ok next"	"next"	"k next step"	"next"	"next step"	"next step"	"next"	"uh next"	"next" + "step" 8 out of 8
		Proposal 5	"k next"	"k next"	"k next step"	"next"	"next step"	"next step"	"next"	"uh next"	[ok] + "next" + "step" 5 out of 8
		Proposal 6	"next"	"k next"	"next step"	"ok next"	"next step"	"last step go back" "next step"	"next step"	"uh next"	[ok] + "next" + "step" 6 out of 8
		Proposal 7	"next"	"next"	"yeah next step"	"next"	"next step"	"next step"	"next"	"ok next"	"next" + "step" 8 out of 8
	Task's Emerging Command										"next" + "step"
	Fixing a Boombox	Proposal 1	"k tim go to the next step"	"uh next"	"k next step"	"next"	"hey google i wanna fix this boombox"	"next"	"go to the next one"	"next"	"next" = "step" 8 out of 8
		Proposal 2	"next step tim"	"next"	"next step"	"next"	"what's the next step"	"next"	"go to the next one"	"next"	"next" + "step" 8 out of 8
		Proposal 3	"next step"	"next"	"next step"	"next"	"what's the next step"	"next"	"go to the next one"	"k next"	"next" + "step" 8 out of 8
		Proposal 4	"next step tim"	"next"	"next step"	"next"	"what's next"	"next"	"go to the next one"	"done next"	"next" 8 out of 8
		Proposal 5	"next step tim"	"next"	"next step"	"next"	"what's next"	"next"	"next"	"next"	"next" 8 out of 8
		Proposal 6	"next step tim"	"next"	"next step"	"next"	"ok what's next"	"next"	"next"	"ok next"	"next" 8 out of 8
		Proposal 7	"next step tim"	"next"	"next step"	"next"	"done that what's next"	"back" "next"	"next"	"yeah next"	"next" 8 out of 8
		Proposal 8	"next"	"next"	"next step"	"next"	"what's next"	"next"	"next"	"yeah done next"	"next" 8 out of 8
	Task's Emerging Command										"next"
	Final Command - based on all proposals (& Number of Users Who Ever Proposed It)										"next" 16 out of 16
	Agreement Rate										1

NOTE: "ok" was not meant as a part of the command, but since it emerged amongst multiple participants, it was analyzed as well. This is further discussed in chapter 4

Figure 75: "step" is used as a part of the cooking a recipe task's emerging command. However, it is not used as a part of the fixing a boombox task's emerging command. Nonetheless, "next" is a common factor in both. Thus, "step" is removed from the final command, and "next" is considered as the final command.

Referent	Most Popular Gesture agreement rate	Voice Command Template for Each Referent (Order of Components Does Not Matter) agreement rate
Interacting with Menu System		
<i>Expand Menu</i>	tap 0.53	"click"/"set"/ other verb ("hit"/"open"/"select"/"program") + object ("button"/"timer") 0.67
<i>Pick Button</i>	tap 0.67	exact words on button 1
<i>Set Slider</i>	drag 1	slider name + value 1
<i>Set Toggle</i>	tap 0.71	toggle value 1
<i>Collapse Menu</i>	tap 0.29 *low agreement rate	"collapse"/ other verb ("erase"/"click") + "menu" 0.86
Environmental Control		
<i>Pick Directional Button/ Specify Direction</i>	tap 0.53	"rotate"/"change"/ other verb ("spin"/"move"/"set"/"select") + object ("blinds"/ UI words for temperature) + value ("left"/"right"/numeric value) 0.87
<i>Open Blinds Entirely</i>	swipe 0.71	"open" + "blinds" 1
<i>Close the Control Panel</i>	tap 0.86	"close" + "menu"/"it" 0.57
Media Control		
<i>Select Media to Play</i>	tap 0.6	"play" + media file name 0.93
<i>Select Button/Modify Playing Status</i>	tap 0.67	"play"/"pause"/ "stop" 0.80
<i>Select Physical Display for Media</i>	drag 0.57	"play" + media file name + "on" + display name 0.86
Following a Workflow		
<i>Go to Next/ Previous Step (not enough users used previous)</i>	tap 0.6	"next" 1

Figure 76: Final Referent Table with the Voice Command Set and Gesture Set. A non-abstracted (more detailed version of this table is available in Appendix G.

In terms of voice commands, users generally opted to use the minimal amount of words needed to get the point across. Moreover, when words were available on the interface, users utilized them as a part of the command. These factors along with other study observations and the results of the semi-structured interview are expanded on in chapter 4.

3.2.2 User Preference

At the end of the study, users were asked which method of control they prefer. This was done to determine whether there was one method that was overwhelmingly preferred. Users were also interviewed on their motivation of picking the preferred method. The preferences are as follows:

Method of Preference	Number of Users
Gestures	7
Voice	6
Mixture: not picking one	3

Figure 77: Users' Method of Preference

There was not enough discrepancy between the 2 methods to declare one as preferable. The reasons for the preferences were as follows (on the following page):

Method of Preference	Reasons Given
Gestures	<ul style="list-style-type: none"> • more subtle with others around/silent • more fun/entertaining • physicality • more immersive • predictability (knew what to do) • simplicity • consistency • seems cooler/can do fancy things with hands • exercise • feels weird to talk like talking to a person • voice commands feel variable • muscle memory/experience • more control/more flexibility in terms of controlling • avoiding language barriers
Voice	<ul style="list-style-type: none"> • more comfortable because not HoloLens user • gestures felt more awkward as maybe not used to it • just more comfortable • less tedious on a big menu /don't want to click continuously on menus • works well for most situations • more fun • better experience • no need of moving hand • just say what you are thinking • more natural- does not feel like you are waiting for it to respond • like to talk • talking is easier • good for laziness • arm tiring out from gesturing on long/detailed menu
Mixture: not picking one	<ul style="list-style-type: none"> • Can do complex, personalized things with gestures while voice is hands free. Also, don't want to set the gaze. • Voice may have problems with accents/dialects. While gestures may be a problem if someone doesn't have five full digits. Prefer voice for cognitive but hand motions for physical analogs (still prefer voice for recipe even though it is physical analog). • Prefer gestures for longer interactions (set up time/temperature) while voice for short interactions/ one word (to fix something)

Figure 78: Reasons given by users for picking the preferred method

After users cited their reasons for their preferences, they were asked whether they would always use their preferred method or if it depends on what is to be done. Everyone said that technology that offers a combination of both voice control and gesture control would be the best.

Chapter 4: Study Observations, User Comments, and Design Implications

Apart from the gesture set and voice commands set, other observations arose out of the study. These were things like recurring or interesting behaviours as well as observations that may offer insight to implementers of AR-controlled IoT technology. Users also provided comments on the entire study experience and preferences. These aspects are discussed here.

4.1 Gesture Round Observations

There were a few recurring aspects in terms of gestures. One aspect was that gestures were noted as being used in all 6 hand orientations with palm facing: up, down, left, right, in, and out. Thus, it should be noted that *AR control technology should incorporate all 6 gesture orientations* to accommodate all users; orientations resulting between these orientations (i.e. half left) should also be considered.

In addition, although all 16 participants were right-handed, 2 participants used their left hand to perform gestures sometimes (even when their right hand was free). Sometimes this was done when the item they wanted to interact with was on the left side of their field-of-view. It can be inferred that the other hand was used due to proximity. Moreover, these people were also switching hands. Thus, it should be noted that *technology should incorporate hand-switching and equal capabilities for both hands*.

Observations were also made regarding how much each finger was being used. The pointer finger was used the most while the thumb was used second most-frequently. Albeit less frequently, the middle, ring, and pinky fingers were also used. An aspect to consider is that fingers may tire out if used too frequently, and the middle, ring, and pinky fingers may tire out before others. Thus,

technology should prioritize the use of the pointer and thumb fingers, but also consider the occasional use of the other three fingers.

Another observation was that there were various types of motion observed within gestures. This consisted of left-to-right motion, right-to-left motion, forward motion, down motion, and up motion. In addition, there was also repetitive motion (like multiple taps, etc.). The aspect to consider is that *technology should incorporate various degrees of freedom while considering gestural motion.*

Also, some gestures were using the entire arm instead of just a hand. An example of this is when something was physically higher up in the field of view, the entire arm was used to reach above. Another example is when the “TV” was far away, and the users used their entire arm to “send” the holographic video thumbnail to it. Gestures remain dependent on the hand, but the arm is used to stretch “reach” of the gesture i.e. if two objects are close, it may require the drag of just the hand whereas two faraway objects might require the drag of the entire arm. Thus, *technology should consider that an entire arm may be used for input.*

Moreover, the fact that users “send” the video thumbnail to the “TV” demonstrates that that when a holographic item needs to interact with a physical object that is far away, participants oftentimes choose to use their gestures in a way that suggests bridging the gap between them. One participant described chosen hand gestures as:

“basically what you see in science fiction and movies”

Thus, this interaction was considered as “cool” by some of the participants, and it aligns with Peshkova et al. 's **instrumented** mental model in which users perform control with an imaginary

tool (i.e. superpowers) [92]. These factors contribute that *gestures should be designed as a means to bridging physical gaps if they exist.*

Also, users oftentimes propose gestures that are influenced by the physical world and how they interact with it. One example was the blinds. When asked to rotate the blinds, some users were rotating an imaginary stick like the one that is often used to physically rotate blinds. When asked to open them entirely, some users gestured as if they are drawing curtains. Thus, it should be noted that *gestures may be influenced by physical world influences.*

4.2 Voice Round Observations

One notable aspect regarding voice commands is that a single **word** was used the most as a command. The second most frequent command type was a **phrase** followed by a complete **sentence**. Thus, such *technology should allow for minimalistic commands.*

In terms of the command content, users chose to: use verbs as a part of their commands, refer to user interface components (i.e. “button”), read words off the user interface, and refer to physical objects. Thus, the *commands should be implemented in alignment with the design of the interface and the anticipated physical environment in mind.*

Also, two participants used names for the system when using voice commands. One participant made up a name (“Hey Tim”) while the other participant used both “Hey Siri” and “Hey Google”. Both users were using voice commands to control a device at least once a week,

so this may be a learned behaviour/ legacy bias. This may be something that may or may not want to be avoided by implementers of such technology. Consequently, this *technology should be designed with the consideration of both: natural conversation structure and previously learned command styles.*

While performing the *following a workflow* scenario (cooking/ fixing a boombox), some users were using an extra word to indicate (often to themselves) that they had Finished the Step. This word came prior to them saying the official command and was often attached to the command itself. Examples of such are “ok next” or “done next” where “next” was the official command they were using to go to the next step of the workflow. This was observed as the user perhaps expecting the system to react to the words “ok” or “done”. Thus, *voice commands should be designed with the consideration of natural behaviour and intuition perhaps even instead of the voice commands that users propose.*

Not enough participants used the **previous step** feature in the *following a workflow* tasks so that it could be included in the consensus set. All 3 voice participants who did use this feature used the word “back”. Thus, *it could be inferred that “back” is a command that could align well with user intuition to go to the previous step.*

4.3 Overall Observations

Over the course of both the voice and gesture rounds, it was noted that users sometimes tended to use multimodal interaction. When asked to perform a voice command, users sometimes supplemented those with gestures and vice versa. This appeared to have been occurring naturally/ subconsciously. An example of this being, a user that was asked to use a voice command to open

the blinds, the user said a command while simultaneously “drawing curtains with their hands”. Thus, it should be noted that *technology should be receptive to multimodal input*.

Also, while people were kneading (and their hands got dirty), three people opted to use a paper towel to clean their hands before moving on to the next step. One was using gestures and two were using voice. The user that was using gestures opted to keep using the same hand to perform gestures after their hand was clean. However, most other gesture users started gesturing with their left hand after their right hand became dirty (and they hadn’t wiped it clean). Thus, *it should be noted that interaction with technology may change based on affordances such as whether the users’ hands are clean*.

4.4 User Comments

4.4.1 Interaction and Strategy

Users discussed their decisions and experience during the interview. One thing that came across various times was that gestures and voice commands are sometimes inspired by interaction that users have from using other technology such as touch screens and computers. Some user quotations below refer to the same:

“I often referred back to what terminology is used in other computer programs...next/previous”

“For music, I actually picked clicking on the icons because that’s what I do on my iPhone...it looks very familiar to the iTunes”

The high popularity of the *tap* gesture could be a direct example of translation of interaction with current touch screens. The same results may not have emerged if this study was performed before

the time of smart phones. Thus, it should be noted that *current technological influences affect how people choose to interact with technology.*

A few users commented on the design aspect of controls that require the neck to be craned up. In the light schedule example, the icon to expand the menu was on the ceiling right next to the light. However, when the user expands the menu, the expanded portion of the menu was at head level. The users indicated that there should be some visual cue informing them that the rest of the menu is below. This is the case since the field of view only lets the user see either above them or in front of them: not both simultaneously. Some users expressed this in the following ways:

“It was hard to see the menu sometimes...I had to move my head to see where things are.”

“It definitely would help the UI to have sort of like a dialogue box so you can see where it’s linking up to...when I clicked on the clock, I wouldn’t have looked down if it wasn’t for your prompt...so some sort of glow to show that what you need is below...like the bottom part of my vision might flash for a bit”

Thus, it should be noted that *when consecutive components are at a distance from one another, some visual cues like arrows, animations, or a glow should indicate where to look next.*

Another emergent discussion was regarding strategy. There was recurring suggestion from the users that they maintained consistency in their strategy throughout the process. An example is that for the *following a workflow* tasks, there were users who would say “next” and others who would say “next step” to go to the next instruction. However, each person stuck to their method. On the other hand, one user proposed three commands to close the video: “pause”, “close the video”, and “shut it off”. The user said that one day they want to use one command while another day they want to use a different one and the system should always work. Also, another user started off by decrementing the thermostat temperature by repeatedly saying “minus” and they eventually

said, “minus 3”. This was a demonstration of a change of strategy. However, it may be that once users overcome a learning curve/ find the strategy that they want, most of them will be consistent.

This is what some users had to say regarding their strategy:

“I wouldn’t change my approach the next time...but I’m not going to use the same words...like it should pick up what I mean by the context...it should expect me to say 100 things or maybe 5 things.”

“I didn’t change my strategy”

Thus, it can be noted that *most users cite themselves as interacting consistently; however, some might want to interact spontaneously.*

4.4.2 Preferences

Despite having method preferences, all participants agreed that there are some cases where they would use their non-preferred method. More specifically, users said that the method they would use would depend on the scenario or the situation. Those who otherwise preferred gestures cited voice commands as valuable for: being more natural, efficiency (when hands are busy), being laziness-friendly, driving, shorter tasks, and multitasking. Those who otherwise preferred voice commands cited gestures as valuable for: gaming, tasks with multiple choices (like calling someone from a list of people), following a workflow with visual cues (like at work), when one must be quiet (like at work), not having to memorize object names (but rather just pointing at the object). Users that preferred a mixture right from the beginning went on to say that voice commands are good for: when hands are busy, being hands free, not having to set the gaze (since some users know that currently gestures work in conjunction with gaze). The same three users considered gestures to be good for: adjusting blinds (since it is usually a hand motion and they are

physical objects), menus, inputting data, moving the holographic video to the “TV”. This is what a couple of users had to say about the pros and cons of each method:

“For voice control you had to pick your words more carefully...I preferred gesture because it seemed cooler...I feel it’s pretty much a scenario type thing like if I am driving, obviously voice is the thing to go to...if I’m at home sitting on my couch, gesture would be my thing...but like if I’m feeling lazy I would use voice...cause it’s very scenario based.”

“short interactions like just one-word things like next, stop, play it would be better using voice...but to set up more complex things even hour or temperature or setting up the printer...maybe like if I had to type the printer’s name...gesture would be better.”

Even among users who picked a preferred method, some stated that a combination of methods would be the best. This discussion also highlighted that the two methods can be liked for different things. Some of these users even said that voice commands and gestures should be usable at the same time where both can be used within the same task at any step (multi-modal). For example, using gesture, but then switching to voice once hands become occupied mid-task. One user stated:

“I would love to combine them both...I would prefer if I could choose my strategy throughout the interaction like...yeah now I’m going to use voice because I got my hands busy...even set up the clock...maybe I’m just leaving home with lots of baggage I could just say...hey set up temperature to 15...so yeah it would be great if I could combine both”

There was also mention of preferring a combination due to the freedom of being able to pick. Two users conveyed that although different methods are good for tasks, one method at a time for one

task would be good enough. For example, if multitasking was important or if they were holding grocery bags, voice would be good. One person said:

“If I came in the house with a bunch of groceries and I said lights on...I would prefer that than to have to manually do it with gestures...overall I’d rather use hand gestures except like maybe a few odd circumstances where multitasking was important...one [method] is still good enough...even if there is like minor conveniences like if you’re multi-tasking or you’re not feeling well enough to get up for it...like that’s still easy to work around...I would never be using both [methods] at the same time.”

Basically, the idea here was that both methods are needed, but not needed together. All these *ideas regarding multi-modal interaction and scenario-dependent favouring of methods convey that no one method is all-encompassing and that a combination/ choices are the best option.*

Chapter 5: Limitations & Conclusion

There are two parts to this chapter. The first part touches on the scope and limitations of the study while the second part focuses on the future work and conclusion.

5.1 Scope & Limitations

5.1.1 Scope

The purpose of this study is to elicit voice commands and gestures to control smart devices/ the Internet of Things with augmented reality. The scope of this research was set in the following ways. In terms of platform, it was limited to headset augmented reality with a focus on the Microsoft HoloLens. Since the focus was purely on how users want to interact, there was no probe on how the backend may influence interaction. The breadth of the research was on home-based scenarios. Furthermore, four scenarios of interest were narrowed down: Interacting with a Menu System, Environmental Control, Media Control, and Following a Workflow. Two tasks with distinct contexts were examined for each scenario (to avoid context-dependent conclusions). 16 users participated: most of whom were computer science majors. Based on the scope, conclusions can be applied to relevant home-based IoT scenarios in which control would be administered with headset augmented reality. These conclusions would likely be most applicable to users who engage with technology on a regular basis.

5.1.2 Limitations

The study was restrictive in terms of some factors. One being the prospect of priming participants. *Chan et al.* [22] noted that **priming** with gestures can greatly influence what gestures participants choose to propose. Moreover, pilot study participants also suggested the avoidance of a “practice round” per say: for the same reasons. Thus, this type of traditional priming (with a practice round) was avoided for both voice commands and gestures as this is an elicitation study.

However, there may be a different way that participants can be primed without influencing their actions.

In terms of **rounds**, ideally, each participant would perform each task of the 8 tasks 2 times: once with voice and once with gestures. If this was the case, each person would perform 16 times. Due to the limitation of **time** and to **reduce learning effects**, participants were asked to perform each task only once: 4 of the 8 tasks with voice, and the other 4 with gestures. However, since there were 4 scenarios: Interacting with a Menu System; Media Control; Environmental Control; and Following a Workflow, it was guaranteed that each participant would complete one task with gesture and the other with voice commands for each scenario. However, it may have been better if each task and method combination could be performed by everyone as that may provide participants with a clearer comparison between the two control methods i.e., they would cook with both gesture and voice, etc. Nonetheless, reducing learning effects could balance this advantage.

Also, the **prompts were read off** to the participants, and that may have influenced the words they proposed for the voice commands. It may have also influenced the length of the commands i.e.) word/ phrase/ sentence: especially for the Interacting with a Menu System Scenario. Other ways of delivering information to the participants may be an option. This may be either them reading the instructions or viewing images or a video.

Moreover, the choice of **words written on the buttons** or other user interface components may influence users in the words that propose in their voice commands. This may have been reduced by using more images instead of words on the UI. On the other hand, it can be seen as beneficial for designing a voice interface.

Another aspect is that **study participants** were recruited via snowball sampling, so they were predominantly from a computer science background. This conveys that the results could be limited to users with similar backgrounds. The results may or may not be different if users from a different **background** (i.e. education majors) were recruited. Also, most participants were in their early 20's, and the results may have been different if the study was conducted with participants that were predominantly in another **age group**. We did not have a random sample, and this impacts generalizability in that the results can be declared as applicable to a certain population rather than everyone. Moreover, results may have been different if there were more participants, but due to the limitation of time, the **number of participants** were limited. In general terms, a larger sample size may have reduced the margin of error.

In addition, participants may or may not interact with technology differently if someone is watching them. This could be the case since some participants discussed using gestures or voice commands based on if they were by themselves or in the **presence of others**. There was also some discussion on how it feels to talk to a machine. Thus, factors such as being video-captured or performing in a research environment with a researcher watching may have influenced the users' interaction.

The **proposal structure** was such that in some cases, users could make multiple proposals for each referent. An example of this was the following a workflow tasks where participants went to the next step 7 or 8 times. This allowed participants to come up with that many varying proposals to accomplish the same output if they so desired. However, there were other cases where users only made one proposal. An example of this was when they collapsed the print queue menu. In this case, users only proposed a single type of input. This is inconsistent as compared to the following a workflow task. The probability of agreement amongst participants changes between

these two examples. We also make it harder to get a high agreement rate if only one proposal is made. Additionally, when comparing two tasks that have the same referent, one task may involve participants making a smaller amount of proposals while the other task may involve participants making more proposals for the same referent. This is also inconsistent as the results of these two tasks are combined and an uneven amount of proposals between the two tasks may cause one task to hold a higher influence in the final result. Examining these factors, it can be said that a more consistent proposal structure may have been more beneficial. However, the tasks were created in an organic way to be as close to real life scenarios as possible. This caused some activities to have multiple proposals of the same thing (going to the next step repeatedly) and other activities to have a single proposal for something (collapsing a menu). Altering task structure for consistency was a possibility and users could have simply been asked to provide multiple proposals for one thing, but either of these options would have come with compromises in the authenticity of the flow of the task at hand. Sticking to the natural occurrence of input in each task provided insight into authentic user interaction styles without much prompting.

5.2 Future Work & Conclusion

5.2.1 Future Work

The scope and limitations highlight gaps that can be filled through future work. For one thing, with more time and a bigger participant pool, other realms could be explored. A study with a group of people from a **different professional or cultural background** may show different results in terms of interaction and preferences. Moreover, simply acquiring **random sampling** or various **age groups** could lead to some variation in results. Small number and non-random sampling could mean that the conclusions are only applicable to a certain group of people i.e. those that are daily users of technology (and go to the University of Calgary, etc.). This is a limitation

that hinders a conclusion that may be applicable to a bigger/ more general pool of potential users. Thus, a bigger and more random sample of participants could provide benefits in future work.

The **user interface** was designed via the WIMP paradigm standard (windows, icons, menus, pointer) [62]. This takes inspiration from current technology, and it is something that a future AR environment could look like. This design standard could have steered interaction such that users mostly tapped. This design provided an easy and consistent way for the users to interact. A relatively consistent user environment was provided where both implementation and interaction could be simple. Alternatively, a different UI style could be explored to see whether and how much that affects results. For example, a UI that avoids holographic buttons altogether and focusses on holographically highlighting physical objects i.e. an orange holographic glow on top of physical objects that hints at a sequence of interactions, could provide entirely different results. For instance, the current UI style where *words written on the UI* influenced voice commands, but some proposed commands could be entirely different if there were no words available on the user interface. A different UI style could also mean that users' method of preference changes. In this way, multiple UI interface styles could be probed on for input interaction variations that may even influence input method preference.

Additionally, **situational and environmental dependencies** could be explored. For example, how do people interact when it is dark or when they are in a room full of people versus by themselves. Perhaps there would be one method that would stand out from the other if certain circumstances are tested. For example, it may be that all users prefer voice commands when it is dark. Furthermore, all participants may prefer gesture when they are at work, etc. Situational and

environmental dependency is a major factor that was discussed by the participants. Thus, this is something that should be explored further, and it may provide deeper insight into the topic.

Furthermore, another prospect that can be examined is **alternative elicitation methods**. For example, what gestures or voice commands do users pick when provided with a list of options? That would provide insight into a set of predetermined gesture and voice command set. Alternatively, users could have been prompted for three proposals to conduct the same step and then asked to pick their favourite. However, the study was conducted in a way such that users were being observed for their first, intuitive proposal. However, other elicitation methods could be explored to see if alternative insights arise.

Another aspect to consider is how interactions would be different if the system was actually **implemented**. Are there other interaction difficulties to consider that may have been overseen because the gulf of execution was removed? There could be things like delays in system response or the system's ability to recognize input based on environment that could encourage the user to favour one input method over another. Thus, although these are the reasons why the experiment was carried out without implementation, it would also be informative to see how technological thresholds could alter user preferences.

Additionally, the experiment could be carried out in a real home **environment** along with **devices** that are actually IoT-embedded. The lab was set up with home scenarios and hardware *was* used. However, it would be interesting and perhaps more realistic for the participants if there were real blinds that were smart instead of a picture of the blinds on a makeshift window. Also, maybe users would feel more at home if they were operating things from a couch instead of an office chair (imaginary couch). The current setup was able to provide insight because the study

was input-based rather than output-based and thus allowed for wizard-of-oz style output. However, a more realistic environment and devices would allow for **environmental affordances** to also become highlighted.

Lastly, other **variables and capabilities** could also be tested. One is conducting the study with another form of augmented reality (instead of headset AR). Another example is probing on other forms of input such as gaze. Another is controlling remotely. How can one control the lighting from another room using voice commands or gesture? Does this add to the input (i.e. “turn on the light” versus “turn on the first light in the garage”)? Another thing to examine is collaborative control. What is the most preferred input method when a family is co-controlling the smart refrigerator (i.e. “add person A’s favourite food to the shopping list”)? The purpose of this research is to contribute in making home life easier for people via IoT and AR technology. Thus, there are virtually unlimited prospects that could be examined in attempt to narrow down on what users want.

5.2.2 Conclusion

This thesis overcomes (in part) the lack of user involvement in the design of gestures and voice commands and presents an exploration of user preferences through an elicitation study. Furthermore, it brings forth the differences between gestural input and voice command input from the perspective of the user.

This topic was narrowed down due to an interest in the convergence of the virtual world and the physical world and how technology can help operate the physical world. A further probe was conducted on augmented reality controls for physical objects. This provided motivation into how IoT devices can be effectively controlled. Furthermore, a literature review was conducted to

find gaps in research. The literature review spanned over: augmented reality hardware and research (especially in relation to headset AR), IoT hardware and research (especially in relation to augmented reality), gestural input and elicitations (as it pertains to current work), voice input and elicitations (as it pertains to current work).

The review provided insight into what needs to be elicited and how it should be elicited, and thus a procedure was developed. The study was conducted with 16 participants who were video-captured while they performed the tasks and gave the interview. The data was analyzed via the open-coding approach. This provided a way to not only observe what the users pick as gestures and voice commands, but also witness the situational factors that influence those choices. Further analysis was performed by deriving observational conclusions regarding the participants' interaction and behavior (i.e. which hand was used when, etc.). A major contribution of this thesis is a novel way to analyze voice commands: *voice command pattern template*. This approach provides a command that would work for most users. Overall, this research provides a better idea of what users want and need in terms of gestures and voice control as well as what their preferences are. This research contributes to bridging the gap between technology and the virtual environment in a way that progresses the human experience.

References

- [1] Acer. 2018. Bringing AR into the classroom: ARCore is now available on Acer Chromebook Tab 10. Retrieved: September 13, 2019 from <https://eu-acerforeducation.acer.com/innovative-technologies/bringing-ar-into-the-classroom-arcore-is-now-available-on-acer-chromebook-tab-10/>.
- [2] Diptanshu Agrawal, S. B. Mane, Apoorva Pacharne, Snehal Tiwari. 2018. IoT Based Augmented Reality System of Human Heart: An Android Application. In *Proceedings of the 2nd IEEE International Conference on International Conference on Trends in Electronics and Informatics (ICOEI '18)*.
- [3] Md Fasiul Alam, Serafeim Katsikas, Olga Beltramelloc, Stathes Hadjiefthymiades. 2017. Augmented and virtual reality based monitoring and safety system: A prototype IoT platform. *Journal of Network and Computer Applications*. 89 (July 1, 2017) 109-119.
- [4] Günter Alce, Maximilian Roszko, Henrik Edlund, Sandra Olsson, Johan Svedberg, Mattias Wallergård. 2017. [POSTER] AR as a User Interface for The Internet of Things—Comparing Three Interaction Models. In *Proceedings of the 2017 IEEE International Symposium on Mixed and Augmented Reality Workshops (ISMARW '17)*.
- [5] Jeremiah Alexander. 2018. The Essential Guide to Mobile AR Gestures. Retrieved: September 13, 2019 from <https://medium.com/inborn-experience/the-essential-guide-to-mobile-ar-gestures-51906df56d3d>.
- [6] AnyMotion. Bloom Gesture. 2019. Retrieved: September 13, 2019 from <https://anymotion.com/en/wissensgrundlagen/ar-brillen/microsoft-hololens>.
- [7] Apple Developer. 2019. Augmented Reality. Retrieved: September 13, 2019 from <https://developer.apple.com/augmented-reality/>.

- [8] AR Critic. 2018. What are the Magic Leap One Input Methods?. Retrieved: September 13, 2019 from <https://arcritic.com/2293/what-are-the-magic-leap-oneinput-methods/>.
- [9] Augmented Reality | Definition of Augmented Reality by Lexico. 2019. Retrieved: September 13, 2019 from https://www.lexico.com/en/definition/augmented_reality.
- [10] Steve Aukstakalnis. 2016. *Practical Augmented Reality: A Guide to the Technologies, Applications, and Human Factors for AR and VR*. Addison-Wesley Professional.
- [11] N. Sarat Chandra Babu. 2017. Keynote 1: Internet of Things (IoT) and Augmented Reality for e-Learning. In *Proceedings of the 5th National Conference on E-Learning & E-Learning Technologies (ELELTECH '17)*.
- [12] Patrick Bader, Huy Viet Le, Julian Strotzer, Niels Henze. 2017. Exploring Interactions with Smart Windows for Sunlight Control. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)*, 2373-2380.
- [13] Abdessamad Badouch, Salah-ddine Krit, Mustapha Kabrane, Khaoula Karimi. 2018. Augmented Reality services implemented within Smart Cities, based on an Internet of Things Infrastructure, Concepts and Challenges: an overview. In *Proceedings of the Fourth International Conference on Engineering & MIS 2018 (ICEMIS '18)*.
- [14] Sandhya Baskaran, Hari Kumar Nagabushanam. 2018. Relational localization based Augmented reality Interface for IOT applications. In *Proceedings of the International Conference on Information and Communication Technology Convergence (ICTC '18)*.
- [15] Ceylan Beşevli, Oğuz Turan Buruk, Merve Erkaya, and Oğuzhan Özcan. 2018. Investigating the Effects of Legacy Bias. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility (DIS '18)*. ACM Press, New York, NY, 277-281.

- [16] Best Buy. 2019. Click and Grow Smart Indoor Garden (SGS8US) - with Basil Seed Capsule Refill - 3 Pack - Dark Grey. Retrieved: September 13, 2019 from https://www.bestbuy.ca/en-ca/product/click-and-grow-smart-indoor-garden-sgs8us-with-basil-seed-capsule-refill-3-pack-dark-grey/10660704?&cmp=knc-s-71700000055105959&gclid=EAIaIQobChMIgsHr6e254wIVBtRkCh1jWg8KEAYYBSABEgKvBfD_BwE&gclsrc=aw.ds.
- [17] Bose. 2019. Bose Home Speaker 500 with Alexa voice control built-in, Black. Retrieved: September 13, 2019 from <https://www.amazon.com/Bose-Speaker-Alexa-control-built/dp/B07FDF9B46>.
- [18] Zac Bowden. 2019. Microsoft HoloLens 2: Everything we know so far. Retrieved: September 13, 2019 from <https://www.windowcentral.com/microsoft-hololens-2-everything-we-know-so-far>.
- [19] BW Online Bureau. 2017. Press the Mic and Trade Your Stock with Voice Commands from this App. Retrieved: September 13, 2019 from <http://bwdisrupt.businessworld.in/article/Press-the-Mic-and-Trade-Your-Stock-with-Voice-Commands-from-this-App/24-05-2017-118838/>.
- [20] Francesco Cafaro. 2006. Using Framed Guessability to Design Gesture Suites for Embodied Interaction. Master's Thesis. University of Illinois, Chicago, USA.
- [21] Joel Carneiro, Rosaldo J. F. Rossetti, Daniel C. Silva, Eugénio C. Oliveira. 2018. BIM, GIS, IoT, and AR/VR Integration for Smart Maintenance and Management of Road Networks: A Review. In *Proceedings of the 2018 IEEE International Smart Cities Conference (ISC2 '18)*.

- [22] Edwin Chan. 2017. *User-defined Single-hand Microgestures*. Master's thesis. University of Calgary, Calgary, Canada.
- [23] Edwin Chan, Frank Maurer. 2016. User Elicitation on Single-hand Microgestures. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16). ACM Press, New York, NY, 34033414.
- [24] Yosun Chang. 2018. Augmented Reality Interfaces for the Internet of Things: Extended Abstract. In *Proceedings of the 2018 ACM SIGGRAPH 2018 Appy Hour* (SIGGRAPH '18).
- [25] David Chaves-Díeguez, Alexandre Pellitero-Rivero, Daniel García-Coego, Francisco Javier González-Castaño, Pedro S. Rodríguez-Hernández, Óscar Piñeiro-Gómez, Felipe J. Gil-Castiñeira, Enrique Costa-Montenegro. 2015. Providing IoT Services in Smart Cities through Dynamic Augmented Reality Markers. *Sensors*. (2015).
- [26] Qing Chen. 2018. *Immersive Analytics Interaction: User Preferences and Agreements by Task Type*. Master's thesis. University of Calgary, Calgary, Canada.
- [27] Sasha Chen, Mei Wang, Xuebin Qin. 2018. Object Recognition of Environmental Information in the Internet of Things Based on Augmented Reality. *MATEC Web of Conferences*. 173, (2018), 02023.
- [28] Albert Choi. 2017. Concept on how Apple can push Augmented Reality and Siri together with iOS 12. Retrieved: September 13, 2019 from <https://medium.com/@AlbertChoi92/how-apple-can-push-augmented-reality-and-siri-together-with-ios-12-94eaf91c5>.

- [29] Alex Colon, Eric Griffith. 2019. The Best Smart Home Devices for 2019. Retrieved: September 13, 2019 from <https://www.pcmag.com/article/303814/the-best-smart-home-devices-for-2019>.
- [30] CTL. 2019. CTL Chromebox CTL22CBX1T 2-in-1 with IP2154T 22" Touch Display. Retrieved: September 13, 2019 from <https://ctl.net/products/ctl-chromebox-ctl22cbx1t-2-in-1-with-ip2154t-22-touch-display>.
- [31] Currys PC World. 2019. Washing Machines. Retrieved September 13, 2019 from https://www.currys.co.uk/gbuk/listing/332_3119_30206_xx_xx/xx-criteria.html.
- [32] Jesse Damiani. 2017. Houzz Launches Enhanced ARKit App to Let You Live Your Dream Home. Retrieved: September 13, 2019 from <https://vrscout.com/news/houzz-virtual-furniture-arkit-app/>.
- [33] Lavinia Andreea Danielescu. 2019. Discoverable Free Space Gesture Sets for Walk-up-and-use Interactions. Doctor of Philosophy Thesis. Arizona State University, Phoenix, USA.
- [34] Haiwei Dong, Nadia Figueroa. 2015. An Elicitation Study on Gesture Attitudes and Preferences Towards an Interactive Hand-Gesture Vocabulary. In *Proceedings of the 23rd ACM international conference on Multimedia (MM '15)*, 999-1002.
- [35] Anthony Elio. 2019. DreamGlass: Adding AR to Everything from Architecture to Education. Retrieved: September 13, 2019 from <https://restechtoday.com/dreamglass-adding-ar-to-everything-from-architecture-to-education/>.
- [36] Epson. 2019. MOVERIO BT-300. Retrieved: September 13, 2019 from <https://www.epson.eu/products/see-through-mobile-viewer/moverio-bt-300>.
- [37] Epson. 2019. MOVERIO Pro BT-2000/2200 Technical FAQ. Retrieved: September 13, 2019 from <https://tech.moverio.epson.com/en/bt-2000/faq.html>.

- [38] Hessam Jahani Fariman, Hasan J. Alyamani, Manolya Kavakli, Len Hamey. 2016. Designing a User-Defined Gesture Vocabulary for an In-Vehicle Climate Control System. In *Proceedings of the 28th Australian Conference on Computer-Human Interaction* (ozCHI '16), 391-395.
- [39] Yasmin Felberbaum, Joel Lanir. 2016. Step by Step: Investigating Foot Gesture Interaction. In *Proceedings of the International Working Conference on Advanced Visual Interfaces* (AVI '16), 306-307.
- [40] Sanjay Ghosh, Jatin Pherwani. 2015. Enabling Naturalness and Humanness in Mobile Voice Assistants. In *Proceedings of IFIP Conference on Human-Computer Interaction* (INTERACT '15).
- [41] Glebys Gonzalez, Naveen Madapana, Rahul Taneja, Lingsong Zhang, Richard Rodgers, Juan P. Wachs. 2018. Looking beyond the gesture: Vocabulary acceptability criteria for gesture elicitation studies. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 62(1), 997-1001.
- [42] Google LLC. 2019. YouTube: Watch, Listen, Stream. Retrieved: September 13, 2019 from <https://apps.apple.com/ie/app/youtube-watch-listen-stream/id544007664>.
- [43] Janet Groeber. 2015. Mixed Reality. *Design : Retail*. 27, 4 (Apr/May 2015) 22.
- [44] Alaric Hamacher, Jahanzeb Hafeez, Roland Csizmazia. 2019. Augmented Reality User Interface Evaluation – Performance Measurement of Hololens, Moverio and Mouse Input. *International Journal of Interactive Mobile Technologies* (iJIM). 13, 03 (2019), 95-107.
- [45] Chris Harrison. 2019. Skinput: Appropriating the Body as an Input Surface. Retrieved: September 13, 2019 from <http://www.chrisharrison.net/index.php/Research/Skinput>.

- [46] Scott Hayden. 2019. Facebook is Working on a Voice Assistant for Oculus Devices, Portal & More. Retrieved: September 13, 2019 from <https://www.roadtovr.com/facebook-working-voice-assistant-oculus-devices-portal/>.
- [47] Alexander G. Hauptmann, Paul McAvinney. 1993. Gestures with Speech for Graphic Manipulation. *International Journal of Man-Machine Studies*. 38, 2 (1993), 231-249.
- [48] Help – Google Glass. 2019. Retrieved: September 13, 2019 from <https://www.google.com/glass/help/>.
- [49] HSO group. 2017. Recreate Hololens Elevator. Retrieved: September 13, 2019 from <https://www.youtube.com/watch?v=2aagEVBzqAg>.
- [50] Valentin Huen. 2019. #78 Reality Editor. *Mit Media Lab*. Retrieved: September 13, 2019 from <https://www.media.mit.edu/videos/labcast-78-reality-editor/?autoplay=true>.
- [51] Edwin L. Hutchins, James D. Hollan, Donald A. Norman. 1985. Direct Manipulation Interfaces. *Human–Computer Interaction*. 1, 4 (1985), 311-338.
- [52] Helge Hüttenrauch, Mike Tate, Martin Böcker, Rosemary Orr, and Françoise Petersen. 2008. Standardising SPOKEN commands for mobile devices and services. In *Proceedings of the International Conference on Mobile Technology, Applications, and Systems (Mobility '08)*.
- [53] IFTTT. 2019. Retrieved: September 13, 2019 from <https://play.google.com/store/apps/details?id=com.ifttt.ifttt&hl=en>.
- [54] Nathan Ingraham. 2016. Augmented Reality Hardware is Still Way Ahead of its Software. Retrieved: September 13, 2019 from <https://www.engadget.com/2016/06/02/augmented-reality-hardware-and-software-still-needs-work/>.

- [55] Internet of things | Definition of Internet of things by Lexico. 2019. Retrieved: September 13, 2019 from https://www.lexico.com/en/definition/internet_of_things.
- [56] iRobot. 2019. Roomba Vacuums. Retrieved: September 13, 2019 from <https://www.irobot.com/roomba>.
- [57] Dongsik Jo. 2019. IoT + AR: pervasive and augmented environments for “Digi-log” shopping experience. *Human-Centric Computing and Information Sciences*. 9, 1 (2019) 1-17.
- [58] Dongsik Jo, Gerard Jounghyun Kim. 2016. ARIoT: Scalable Augmented Reality Framework for Interacting with Internet of Things Appliances Everywhere. In *Proceedings of the 2016 IEEE Transactions on Consumer Electronics*.
- [59] Seb Joseph. 2017. How Ikea is Using Augmented Reality. Retrieved: September 13, 2019 from <https://digiday.com/marketing/ikea-using-augmented-reality/>.
- [60] John A. Karasinski, Richard Joyce, Colleen Carroll, Jack Gale, Steven Hillenius. 2017. An Augmented Reality/Internet of Things Prototype for Just-in-time Astronaut Training. In *Proceedings of the International Conference on Virtual, Augmented, and Mixed Reality (VAMR '17)*, 248-260.
- [61] Sanshzar Kettebekov, Rajeev Sharma. 2001. Toward natural gesture/speech control of a large display. *Lecture Notes in Computer Science*. (2001), 221-234.
- [62] Stephen Kimani. 2009. WIMP Interfaces. In: *Ling Liu, M. Tamer Özsu (Eds.) Encyclopedia of Database Systems*. (2009), 3529-3533.
- [63] G. Koutitas, J. Jabez, C. Grohman, C. Radhakrishna, V. Siddaraju, S. Jadon. 2018. Demo/poster abstract: XReality research lab — Augmented reality meets Internet of

- Things. In *Proceedings of IEEE Conference on Computer Communications Workshops, INFOCOM Wksp*s (INFOCOM '18).
- [64] Ben Lang. 2018. Meta Forced to Furlough After Chinese Lead Backs Out of New Investment. Retrieved: September 10, 2018 from <https://www.roadtovr.com/meta-struggles-to-raise-cash-after-chinese-lead-investor-backs-out/>.
- [65] Teemu Leppänen, Arto Heikkinen, Antti Karhu, Erkki Harjula, Jukka Riekk, Timo Koskela. 2014. Augmented Reality Web Applications with Mobile Agents in the Internet of Things. In *Proceedings of the 8th IEEE International Conference on Next Generation Mobile Applications, Services and Technologies* (NGMAST '14).
- [66] DoYoung Lee, Youryang Lee, Yonghwan Shin, Ian Oakley. 2018. Designing Socially Acceptable Hand-to-Face Input. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology* (UIST '18).
- [67] Sang-Su Lee, Jeonghun Chae, Hyunjeong Kim, Youn-kyung Lim, Kun-pyo Lee. 2013. Towards More Natural Digital Content Manipulation via User Freehand Gestural Interaction in a Living Room. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing* (UbiComp '13), 617-626.
- [68] Nestor Lobo. 2016. Intelli-mirror: An augmented reality based IoT system for clothing and accessory display. In *Proceedings of the International Conference on Internet of Things and Applications* (IOTA '16).
- [69] Daniel Lavorato. 2019. Actually, 'Pokémon Go' Is More Popular Than Ever. Retrieved: September 13, 2019 from <https://junkee.com/pokemon-go-still-thriving/196317>.

- [70] LG Newsroom. 2014. Lg Makes Smart TV Simple with New Webos Smart TV Platform. Retrieved: September 13, 2019 from <http://www.lgnewsroom.com/2014/01/lg-makes-smart-tv-simple-with-new-webos-smart-tv-platform/>.
- [71] Cherlynn Low. 2019. Lenovo launches ThinkReality AR and VR headset for enterprises. Retrieved: September 13, 2019 from <https://www.engadget.com/2019/05/13/lenovothinkreality-ar-vr-headset-hololens-2/#/>.
- [72] M. Makolkina, A. Koucheryavy, A. Paramonov. 2017. The Models of Moving Users and IoT Devices Density Investigation for Augmented Reality Applications. In *Proceedings of the 2017 Conference of the Internet of Things, Smart Spaces, and Next Generation Networks and Systems* (NEW2AN '17).
- [73] mattzmsft, BrandonBray, jessecculloch, rwinj, BrandonBray-zz, mr0ng, thetuvix, grbury. Docs.microsoft.com. 2019. Gestures - Mixed Reality. Retrieved: September 13, 2019 from <https://docs.microsoft.com/en-us/windows/mixed-reality/gestures>.
- [74] Keenan R. May, Thomas M. Gable, Bruce N. Walker. Designing an In-Vehicle Air Gesture Set Using Elicitation Methods. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (AutomotiveUI '17), 74-83.
- [75] Jay McGregor. 2019. Google Home Vs. Amazon Echo: 45 Complex Questions, 1 Clear Winner. *Forbes*. Retrieved: September 13, 2019 from <https://www.forbes.com/sites/jaymcgregor/2018/03/15/google-home-vs-amazon-echo-45-complex-questions-1-clear-winner/#79e9dab8f284>.

- [76] Michael Mimoso. 2016. Skyping and Typing the Latest Threat to Privacy. Retrieved: September 13, 2019 from <https://threatpost.com/skyping-and-typing-the-latest-threat-to-privacy/121387/>.
- [77] Konstantinos Michalakakis, John Aliprantis, George Caridakis. 2018. Visualizing the Internet of Things: Naturalizing Human-Computer Interaction by Incorporating AR Features. *IEEE Consumer Electronics Magazine*. 7, 3 (May 2018), 64-72.
- [78] Meredith Ringel Morris. 2012. Web on the wall. In *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces (ITS '12)*, 95-104.
- [79] Meredith Ringel Morris, Andreea Danielelescu, Steven Drucker, Danyel Fisher, Bongshin Lee, M.C. Schraefel, and Jacob O. Wobbrock. 2014. Reducing legacy bias in gesture elicitation studies. *interactions*. 21, 3 (May/June 2014), 40-45.
- [80] Michael Muller. 2012. Grounded Theory Methods. Retrieved September 13, 2019 from <https://www.slideshare.net/traincroft/muller-and-chua-brainstorming-for-japan-chi-2012>.
- [81] Georgios Mylonas, Christos Triantafyllis, Dimitrios Amaxilatis. 2019. An Augmented Reality Prototype for supporting IoT-based Educational Activities for Energy-efficient School Buildings. *Electronic Notes in Theoretical Computer Science*. 343 (May 4, 2019), 89-101.
- [82] Madduri Narsimlu, Midde Venkateswarlu Naik, Mallojjala Raja Sree. 2017. An innovative research on augmented reality to control status of physical light using MAYA as part of IoT. In *Proceedings of the 2017 IEEE International Conference on Smart Technologies for Smart Nation (SmartTechCon '17)*.

- [83] Michael Nebeling, Alexander Huber, David Ott and Moira C. Norrie. 2014. Web on the Wall Reloaded. In *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces* (ITS '14), 15-24.
- [84] Michael Nebeling, David Ott, Moira C. Norrie. 2015. Kinect analysis: A system for recording, analysing and sharing multimodal interaction elicitation studies. In *Proceedings of the 7th ACM SIGCHI Symposium on Engineering Interactive Computing Systems* (EICS '15), 142-151.
- [85] Mohammad Obaid, Felix Kistler, Gabrielė Kasparavičiūtė, Asim Evren Yantaç, Morten Fjeld. 2016. How Would You Gesture Navigate a Drone?: A User-Centered Approach To Control a Drone. In *Proceedings of the 20th International Academic Mindtrek Conference* (AcedemicMindTrek '16), 113-121.
- [86] Open Connectivity Foundation. 2019. OCF - UPnP Standards & Architecture. Retrieved: September 13, 2019 from <https://openconnectivity.org/developer/specifications/upnp-resources/upnp>.
- [87] Scott Orgera. 2019. The Top 100 Google Assistant and Google Home Commands. Retrieved: September 13, 2019 from <https://www.lifewire.com/top-googleassistant-and-google-home-commands-4158256>.
- [88] Simon Lysander Overstall. 2017. Work in Progress: “I Am Afraid”. Retrieved: September 13, 2019 from <http://www.simonlysander.net/news/2017/5/20/i-am-afraid>.
- [89] Williams Pelegrin. 2019. Phones compatible with ARCore: What are your options?. Retrieved: September 13, 2019 from <https://www.androidauthority.com/compatible-arcore-phones-932844/>.

- [90] Tommy Palladino. 2018. Former Meta Exec Launches DreamGlass AR Headset, Shipping at End of June for \$399. Retrieved: September 13, 2019 from <https://meta.reality.news/news/former-meta-exec-launches-dreamglass-ar-headset-shipping-end-june-for-399-0185131/>.
- [91] Pete Pachal. 2018. How Magic Leap Compares to Microsoft HoloLens. Retrieved: September 13, 2019 from <https://mashable.com/article/magic-leap-vs-microsoft-hololens-hands-on/>.
- [92] Ekaterina Peshkova, Martin Hitz, David Ahlström. 2016. Exploring User-Defined Gestures and Voice Commands to Control an Unmanned Aerial Vehicle. *Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*. (2016), 47-62.
- [93] Pilaiwan Phupattanasilp. 2019. Augmented Reality in the Integrative Internet of Things (AR-IoT): Application for Precision Farming. *Sustainability*. MDPI AG.
- [94] Boris Pokric, Srdjan Krco, Dejan Drajjic, Maja Pokric, Vladimir Rajs, Zivorad Mihajlovic, Petar Knezevic, Dejan Jovanovic. 2015. Augmented Reality Enabled IoT Services for Environmental Monitoring Utilising Serious Gaming Concept. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*.
- [95] Boris Pokrić, Srđan Krčo, Maja Pokrić. 2014. Augmented Reality Based Smart City Services Using Secure IoT Infrastructure. In *Proceedings of the 28th IEEE International Conference on Advanced Information Networking and Applications Workshops (AINAW '14)*.
- [96] Boris Pokrić, Srđan Krčo, Maja Pokrić, Petar Knežević, Dejan Jovanović. 2015. Engaging citizen communities in smart cities using IoT, serious gaming and fast markerless

- Augmented Reality. In *Proceedings of the 2015 International Conference on Recent Advances in Internet of Things (RIoT '15)*.
- [97] Jannish A. Purmaissur, Praveer Towakel, Shivanand P. Guness, Amar Seeam, Xavier A. Bellekens. 2018. Augmented-Reality Computer-Vision Assisted Disaggregated Energy Monitoring and IoT Control Platform. In *Proceedings of the 2018 IEEE Conference on Intelligent and Innovative Computing Applications (ICONIC '18)*.
- [98] Narayanan Raman. 2019. The Smart Home Promise – Power Consumption as a Key Driver. Retrieved: September 13, 2019 from <https://www.imgtec.com/blog/how-low-powered-wi-fi-sensors-are-the-future-of-iot/>.
- [99] Jason Rambach, Alain Pagani, Didier Stricker. 2017. [POSTER] Augmented Things: Enhancing AR Applications leveraging the Internet of Things and Universal 3D Object Tracking. In *Proceedings of the 2017 IEEE International Symposium on Mixed and Augmented Reality Workshops (ISMARW '17)*.
- [100] Vincy Rajan, N.V. Sobhana, R. Jayakrishnan. 2018. Machine Fault Diagnostics and Condition Monitoring Using Augmented Reality and IoT. In *Proceedings of the 2nd IEEE International Conference on Intelligent Computing and Control Systems (ICICCS '18)*.
- [101] Zulqarnain Rashid, Joan Melià-Seguí, Rafael Pous, Enric Peig. 2017. Using Augmented Reality and Internet of Things to improve accessibility of people with motor disabilities in the context of Smart Cities. *Future Generation Computer Systems*. 76. (Nov. 2017). 248-261.
- [102] Roberto Revetria, Flavio Tonelli, Lorenzo Damiani, Melissa Demartini, Federico Bisio, Nicola Peruzzo. 2019. A Real-Time Mechanical Structures Monitoring System Based On

- Digital Twin, Iot and Augmented Reality. In *Proceedings of the IEEE Spring Simulation Conference* (SpringSim '19).
- [103] Isabel Benavente Rodriguez. 2016. *Gesture Elicitation Study on How to Opt-in & Opt-out from Interactions with Large Public Displays*. Master's thesis. University College London, London, UK.
- [104] Janko Roettgers. 2019. What the Oculus Quest Can Teach Us About the Future of Mixed Reality. Retrieved: October 14, 2019 from <https://variety.com/2019/digital/news/oculus-quest-guardian-system-mixed-reality-1203223870/>.
- [105] Marco Romano, Andrea Bellucci, Ignacio Aedo. 2015. Understanding Touch and Motion Gestures for Blind People on Mobile Devices. In *Proceedings of the 2015 IFIP Conference on Human-Computer Interaction* (INTERACT '15), 38-46.
- [106] Jaime Ruiz, Daniel Vogel. 2015. Soft-Constraints to Reduce Legacy and Performance Bias to Elicit Wholebody Gestures with Low Arm Fatigue. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI '15), 3347-3350.
- [107] Gustavo Alberto Rovelo Ruiz, Davy Vanacken, Kris Luyten, Francisco Abad, Emilio Camahort. 2014. Multi-viewer gesture-based interaction for omni-directional video. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems* (CHI '14), 4077-4086.
- [108] Anshel Sag. 2018. The Long Road Ahead for Magic Leap and AR. Retrieved: October 14, 2019 from <https://www.forbes.com/sites/moorinsights/2018/10/03/the-long-road-ahead-for-magic-leap-and-ar/#46b3db912619>.

- [109] Matthias Seuter, Eduardo Rodriguez Macrillante, Gernot Bauer, Christian Kray. 2018. Running with Drones: Desired Services and Control Gestures. In *Proceedings of the 30th Australian Conference on Computer-Human Interaction (ozCHI '18)*.
- [110] Choonsung Shin, Byoung-Ha Park, Gwang-Mo Jung, Sung-Hee Hong. 2014. Mobile Augmented Reality Mashup for Furture IoT Environment. In *Proceedings of the 2014 IEEE 11th Intl Conf on Ubiquitous Intelligence & Computing and 2014 IEEE 11th Intl Conf on Autonomic & Trusted Computing and 2014 IEEE 14th Intl Conf on Scalable Computing and Communications and Associated Symposia/Workshops (UIC-ATC '14)*.
- [111] Small rectangular living room kitchen combo farmhouse style | Country Living Room Decorating Ideas Sloped Ceiling Bedroom Farmhouse ... | kitchen decor | Kitchen family rooms, Living room kitchen, Small cottage interiors. 2019. Retrieved: October 13, 2019 from <https://www.pinterest.ca/pin/419538521520759082/>.
- [112] Sobro. 2019. White/Black Smart Coffee Table (Part number: SOCTB300WHBK). Retrieved September 13, 2019 from <https://www.wayfair.com/furniture/pdx/sobro-smart-coffee-table-sobr1001.html>.
- [113] Anthony Spadafora. 2016. Philips Hue Lights Can Spread IoT Worm. Retrieved September 13, 2019 from <https://betanews.com/2016/11/14/philips-hue-light-bulbs-worm-vulnerable/>.
- [114] Heesuk Son, Seungwook Han, Dongman Lee. 2012. Contextual Information Provision on Augmented Reality with IoT-Based Semantic Communication. In *Proceedings of the 2012 International Symposium on Ubiquitous Virtual Reality (ISUVR '12)*.

- [115] Statista Research Department. 2016. Internet of Things - number of connected devices worldwide 2015-2025. Retrieved: October 14, 2019 from <https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/>.
- [116] Forrest Stroud. 2019. IoT-Internet of Things. Retrieved: September 13, 2019 from https://www.webopedia.com/TERM/I/internet_of_things.html.
- [117] Jim Tanous. 2019. The Lenovo ThinkReality A6 Is a Lightweight AR Headset for Businesses. Retrieved: September 13, 2019 from <https://pcper.com/2019/05/lenovo-thinkreality-a6/>.
- [118] The Ambient. 2019. How to get started with Smart Blinds: Top Brands and Options for Your Shades. Retrieved: September 13, 2019 from <https://www.the-ambient.com/guides/best-smart-blinds-homekit-alexa-448>.
- [119] Radu-Daniel Vatavu, Jacob O. Wobbrock. 2016. Between-Subjects Elicitation Studies. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16), 3390-3402.
- [120] John Voorhees. 2018. Philips Hue App Update Enhances Light Management and Adds 30 New Designer Scenes. Retrieved: October 13, 2019 from <https://www.macstories.net/reviews/philips-hue-app-update-enhances-light-management-and-adds-30-new-designer-scenes/>.
- [121] Dennis Vroegop, James Ashley. 2017. App Development for Microsoft HoloLens. Retrieved: October 13, 2019 from <https://www.lynda.com/Windows-tutorials/App-Development-Microsoft-HoloLens/587658-2.html>.

- [122] Jacob O. Wobbrock, Meredith Ringel Morris, Andrew D. Wilson. 2009. User-Defined Gestures for Surface Computing. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '09)*, 1083-1092.
- [123] Megan Wollerton. 2018. How One Little Thermostat Started a Design Revolution. Retrieved: October 14, 2019 from <https://www.cnet.com/news/how-the-nest-learning-thermostat-started-a-design-revolution/>.
- [124] Huiyue Wu, Yu Wang, Jiali Qiu, Jiayi Liu, Xiaolong (Luke) Zhang. 2018. User-defined gesture interaction for immersive VR shopping applications. *Behaviour & Information Technology*. 38, 2. (Dec. 2018), 726-741.
- [125] Ho-Chih Yu. 2014. The Haptic Feedback Design of Augmented Reality Virtual Keyboard on The Air. In *Proceedings of the 2014 UAHCI International Conference on Universal Access in Human-Computer Interaction (UAHCI '14)*, 463-472.
- [126] Ryan Zehm. 2016. GoogleVR 0.8 Update, Gaze Input, 2000 VR Developers, and more!. Retrieved: October 14, 2019 from <https://www.youtube.com/watch?v=NfG63Ge3aQk>.

Appendix A: Pre-Study Questionnaire

Pre-Study Questionnaire Participant ID : _____ (To be filled in by researcher)

Gender : _____ Age : _____

1. How often do you use remote controls to control a device (controlling from a phone, computer, etc) (for work/non-work related)?

☐ Daily ☐ 2-5 times a week ☐ once a week ☐ once a month ☐ very rarely

2. How often do you use voice commands to control devices (ie. Where you are speaking to control how a specific device behaves- siri, google home, etc)?

☐ Daily ☐ 2-5 times a week ☐ once a week ☐ once a month ☐ very rarely

3. How often do you use augmented reality *without* a headset- ie. On a phone (for work/ nonwork-related)?

☐ Daily ☐ 2-5 times a week ☐ once a week ☐ once a month ☐ very rarely

4. How often do you use augmented reality *with* a headset (for work/ non-work-related)?

☐ Daily ☐ 2-5 times a week ☐ once a week ☐ once a month ☐ very rarely

5. How often do you use virtual reality *with* a headset (for work/ non-work-related)?

☐ Daily ☐ 2-5 times a week ☐ once a week ☐ once a month ☐ very rarely

6. If you have experience with any of the following immersive technologies (i.e. Virtual reality, mixed reality, augmented reality)- on which platforms? (i.e. head set, tablet, phone, etc.) (i.e. snapchat filters, etc.)?

7. By definition, the internet of things (IoT) is: *interconnection via the Internet of computing devices embedded in everyday objects, enabling them to send and receive data*. What experience do you have with IoT, if any? If so, please describe some example of how you have used these technologies. What are some common activities you perform on these?

Appendix B: Part 1 (Gestures first script)

Researcher Script (Including Semi-Structured Interview)

WELCOME [10 minutes]

Welcome! Thank you for your interest in my study. Please have a seat and carefully read over this consent form. Let me know if you have any questions.

Pre-study questionnaire (See Appendix) + Assign Participant ID [Questionnaire Wrangler]

INSTRUCTIONS [5 minutes]

Today I'll be observing you while you interact with IoT home scenarios in which you will control physical objects with AR. You will see menus in the headset that are meant to control the things around you. There will be 8 scenarios in total- separated into 2 rounds. In round 1, I will ask you to interact with 4 scenarios in which you will complete the task using gesture navigation. In round 2, I will ask you to interact with 4 other scenarios in which you will complete the task using voice navigation.

After each of the two rounds, I'll take you aside for an interview. Any questions?

Round 1) For this round, I would like you to use gestures to navigate the AR menu in each of the scenarios. I will also be video-capturing as you perform the task. For each scenario, I will read out a prompt and the included steps, and you can decide what gestures you would want to use to finish the task. You may choose to use gestures to use the menu or use gestures to complete the task without the use of the menu.

[fit them with the headset]

Please let me know if at any point you don't see the AR menu or are uncomfortable with the headset.

Round 2) For this round, I would like you to use voice commands to navigate through the scenarios. For each scenario, I will read out a prompt and the included steps, and you can decide what voice commands you would want to use to finish the task. You may choose to use voice commands to use the menu or use voice commands to complete the task without the use of the menu.

Post-Task Interview Protocol [Completed after each of the 2 rounds] 7 minutes

What was your strategy while navigating the controls?

a. How did you decide what the best (gesture/voice command) was to navigate specific parts of the control system?

How would you change your approach for the next time you interact with a system like that?

Was there anything in the features of the controls that you did not like or anything that you would change that you would like to use?

Can you describe a point where it felt easy (to accomplish the task at hand)?

a. Why did it feel easier?

b. Tell me more

Can you describe a point where it felt difficult (to accomplish the task at hand)?

a. Why did it feel easier?

b. Tell me more

Can you describe an instance in which you felt particularly satisfied?

Were you confused or irritated at any point? Could you describe a point where you felt frustrated?

Did you change your strategy of maneuvering the system at any point? What stood out?

Post-Study Interview Protocol [Comparison (after both tasks)] 10 minutes

How did your strategy for accomplishing the tasks differ between the two rounds with one control round being the gesture navigation and the other round being the voice navigation?

Did you find your ability to complete the tasks at hand improved or got worse across the two methods?

Which method of maneuvering the control system did you prefer?

a. Why did you prefer that method?

b. How would you feel about maneuvering that control system again?

Do you think that there is any advantage or disadvantage to using one control method over another?

a. How so?

b. Were there more pros/cons in one version?

Do you feel like you would use one control method over the other to control devices?

Would you always use one method or the other, or would it depend on what you are using it for?

a. Why?

<p>Printing Scenario</p> <p>In this scenario, you will be connecting your computer to a print queue. Please look at the computer so that the hologram appears in the right position.</p> <p>Do you see an AR button? Let's begin the prompt:</p> <ol style="list-style-type: none">1)Using gesture, manipulate the button so that the next options in the menu appear.2)Connect as input3)Pick print queue4)Pick printer 35)Collapse the menu	<p>Blinds Scenario</p> <p>In this scenario, you will be controlling blinds on your windows. Imagine that this whole display is your window covered in blinds in vertical orientation (as marked by the paper on the display).</p> <p>Please look at the display so that the holographic menu appears in the correct position.</p> <p>Do you see a menu? You may choose to finish the task with or without the use of the menu. So use gestures to either manipulate the menu or use gestures to directly manipulate the blinds. Let's begin the prompt:</p> <ol style="list-style-type: none">1)Rotate blinds to the left2)Rotate them to the right3)Make it so that they are not covering the window any more	<p>Speakers Scenario</p> <p>Please take a standing position. In this scenario, you will be playing music from a speaker system. The locations of the speakers is not specified. They can be imagined as embedded on the ceiling of the room as if they will play when an audio piece is selected.</p> <p>Please look straight ahead while I load the scenario.</p> <p>Do you see an audio thumbnail? You may choose to finish the task with or without the use of the menu. So use gestures to either manipulate the menu or use gestures to directly manipulate the supposed speakers. Let's begin the prompt:</p> <ol style="list-style-type: none">1) Pick the song so that it will play2) Pause the song3) Play the song	<p>Cooking Scenario</p> <p>In this scenario, you will be following instructions to make some dough. The ingredients are on the other table.</p> <p>Please take stand up and look straight ahead while I load the scenario.</p> <p>Do you see a menu? Let's begin.</p> <p>There will be no prompt given to you in this scenario. You will decide when to use gestures to move on to the next step or go back to the previous step if needed. Please follow the workflow until you reach the end of the steps.</p>
<p>Lights Scenario</p> <p>In this scenario, you will be setting a lights schedule for what days and time you want this light to automatically turn on and off.</p> <p>Please look straight ahead so that the hologram appears in the right position when I load it.</p> <p>Please look up- do you see a holographic clock next to the light?</p> <p>Let's begin the prompt:</p> <ol style="list-style-type: none">1)Using voice commands, manipulate the clock so that the next options in the menu appear. Look straight ahead.2)Pick time3)Pick start4)Set hour to 75)Set minutes to 256)Set seconds to 307)Set to AM8)Set as done9)Pick end10) Set hour to 4 Set minutes to 1511) Set seconds to 012) Set to PM13) Pick Day14) Set Monday15) Set Saturday16) Set as done	<p>Thermostat Scenario</p> <p>Thank you. In the next scenario, you will be controlling the building temperature.</p> <p>Please look straight ahead so that the hologram appears in the correct position.</p> <p>Do you see a menu? You may choose to finish the task with or without the use of the menu. So use voice commands to either manipulate the menu or use voice commands to directly manipulate the temperature. Let's begin the prompt.</p> <ol style="list-style-type: none">1) Please change the hold at temperature to 24.2) Now change it to 18.3) Please close the thermostat control	<p>Video Display Scenario</p> <p>Please take a seat. In this scenario, you will be playing a video onto one of two displays. [This] is display 1 and [this is display 2].</p> <p>Please look straight ahead.</p> <p>Do you see a video thumbnail? You may choose to finish the task with or without the use of the menu. So use voice commands to either manipulate the menu or use voice commands to directly manipulate the display that is playing the video. Let's begin the prompt:</p> <ol style="list-style-type: none">1) Make it so that the video will play on display 1 when you decide to play it.2) Pick the video 1 so that it will play3) Close the video player so that it will stop playing	<p>Boombox Scenario</p> <p>In this scenario, you will be following instructions to fix a boombox.</p> <p>Please take a seat and look straight ahead.</p> <p>Do you see a menu?</p> <p>Let's begin. There will be no prompt given to you in this scenario. You will decide when to use voice commands to move on to the next step or go back to the previous step if needed. Please follow the workflow until you reach the end of the steps.</p>

Group 1: Gestures first

Appendix B: Part 2 (Voice first script)

Researcher Script (Including Semi-Structured Interview)

WELCOME (10 minutes)

Welcome! Thank you for your interest in my study. Please have a seat and carefully read over this consent form.. Let me know if you have any questions.

Pre-study questionnaire (See Appendix) + Assign Participant ID [Questionnaire Wrangler]

INSTRUCTIONS (5 minutes)

Today I'll be observing you while you interact with IoT home scenarios in which you will control physical objects with AR. You will see menus in the headset that are meant to control the things around you. There will be 8 scenarios in total- separated into 2 rounds. In round 1, I will ask you to interact with 4 scenarios in which you will complete the task using voice navigation. In round 2, I will ask you to interact with 4 other scenarios in which you will complete the task using gesture navigation.

After each of the two rounds, I'll take you aside for an interview. Any questions?

Round 1) For this round, I would like you to use voice commands to navigate through the scenarios. I will also be video-capturing as you perform the task. For each scenario, I will read out a prompt and the included steps, and you can decide what voice commands you would want to use to finish the task. You may choose to use voice commands to use the menu or use voice commands to complete the task without the use of the menu.

[fit them with the headset]

Please let me know if at any point you don't see the AR menu or are uncomfortable with the headset.

Round 2) For this round, I would like you to use gestures to navigate the AR menu in each of the scenarios. For each scenario, I will read out a prompt and the included steps, and you can decide what gestures you would want to use to finish the task. You may choose to use gestures to use the menu or use gestures to complete the task without the use of the menu.

Post-Task Interview Protocol (Completed after each of the 2 rounds) 7 minutes

1. What was your strategy while navigating the controls?
 - a. How did you decide what the best [gesture/voice command] was to navigate specific parts of the control system?
2. How would you change your approach for the next time you interact with a system like that?
3. Was there anything in the features of the controls that you did not like or anything that you would change that you would like to use?
4. Can you describe a point where it felt easy [to accomplish the task at hand]?
 - a. Why did it feel easier?
 - b. Tell me more
5. Can you describe a point where it felt difficult [to accomplish the task at hand]?
 - a. Why did it feel easier?
 - b. Tell me more
6. Can you describe an instance in which you felt particularly satisfied?
7. Were you confused or irritated at any point? Could you describe a point where you felt frustrated?
8. Did you change your strategy of maneuvering the system at any point?
9. What stood out?

Post-Study Interview Protocol (Comparison (after both tasks)) 10 minutes

1. How did your strategy for accomplishing the tasks differ between the two rounds with one control round being the gesture navigation and the other round being the voice navigation?
2. Did you find your ability to complete the tasks at hand improved or got worse across the two methods?
 - a. Why did you prefer that method?
 - b. How would you feel about maneuvering that control system again?
4. Do you think that there is any advantage or disadvantage to using one control method over another?
 - a. How so?
 - b. Were there more pros/cons in one version?
5. Do you feel like you would use one control method over the other to control devices?
6. Would you always use one method or the other, or would it depend on what you are using it for?
 - a. Why?

Printing Scenario

In this scenario, you will be connecting your computer to a print queue. Please look at the computer so that the hologram appears in the right position.

Do you see an AR button? Let's begin the prompt:

- 1)Using voice commands, manipulate the button so that the next options in the menu appear.
- 2)Connect as input
- 3)Pick print queue
- 4)Pick printer 3
- 5)Collapse the menu

Blinds Scenario

In this scenario, you will be controlling blinds on your windows. Imagine that this whole display is your window covered in blinds in vertical orientation (as marked by the paper on the display).

Please look at the display so that the holographic menu appears in the correct position.

Do you see a menu? You may choose to finish the task with or without the use of the menu. So use voice commands to either manipulate the menu or use voice commands to directly manipulate the blinds. Let's begin the prompt:

- 1)Rotate blinds to the left
- 2)rotate them to the right
- 3)Make it so that they are not covering the window any

Speakers Scenario

Please take a standing position. In this scenario, you will be playing music from a speaker system. The locations of the speakers is not specified. They can be imagined as embedded on the ceiling of the room as if they will play when an audio piece is selected.

Please look straight ahead while I load the scenario.

Do you see an audio thumbnail? You may choose to finish the task with or without the use of the menu. So use voice commands to either manipulate the menu or use voice commands to directly manipulate the supposed speakers. Let's begin the prompt:

- 1) Pick the song so that it will play
- 2) Pause the song
- 3) Play the song

Cooking Scenario

In this scenario, you will be following instructions to make some dough. The ingredients are on the other table.

Please take stand up and look straight ahead while I load the scenario.

Do you see a menu? Let's begin.

There will be no prompt given to you in this scenario. You will decide when to use voice commands to move on to the next step or go back to the previous step if needed. Please follow the workflow until you reach the end of the steps.

Lights Scenario

In this scenario, you will be setting a lights schedule for what days and time you want this light to automatically turn on and off.

Please look straight ahead so that the hologram appears in the right position when I load it.

Please look up- do you see a holographic clock next to the light?

Let's begin the prompt:

- 1)Using gestures, manipulate the clock so that the next options in the menu appear. Look straight ahead.
- 2)Pick time
- 3)Pick start
- 4)Set hour to 7
- 5)Set minutes to 25
- 6)Set seconds to 30
- 7)Set to AM
- 8)Set as done
- 9)Pick end
- 10) Set hour to 4 Set minutes to 15
- 11) Set seconds to 0
- 12) Set to PM
- 13) Pick Day
- 14) Set Monday
- 15) Set Saturday
- 16) Set as done

Thermostat Scenario

Thank you. In the next scenario, you will be controlling the building temperature.

Please look straight ahead so that the hologram appears in the correct position.

Do you see a menu? You may choose to finish the task with or without the use of the menu. So use gestures to either manipulate the menu or use gestures to directly manipulate the temperature. Let's begin the prompt.

- 4)Please change the hold at temperature to 24.
- 5)Now change it to 18.
- 6)Please close the thermostat control

Video Display Scenario

Please take a seat. In this scenario, you will be playing a video onto one of two displays. [This] is display 1 and [this is display 2].

Please look straight ahead.

Do you see a video thumbnail? You may choose to finish the task with or without the use of the menu. So use gestures to either manipulate the menu or use gestures to directly manipulate the display that is playing the video. Let's begin the prompt:

- 4)Make it so that the video will play on display 1 when you decide to play it.
- 5)Pick the video 1 so that it will play
- 6)Close the video player so that it will stop playing

Boombbox Scenario

In this scenario, you will be following instructions to fix a boombbox.

Please take a seat and look straight ahead. Do you see a menu?

Let's begin. There will be no prompt given to you in this scenario. You will decide when to use gestures to move on to the next step or go back to the previous step if needed. Please follow the workflow until you reach the end of the steps.

Group 2: Voice First

Appendix C: Results of Pre-Study Questionnaire

Usage of Remote Controls	Usage of Voice Commands to Control Devices	Usage of Augmented Reality without a Headset	Usage of Augmented Reality with a Headset	Usage of Virtual Reality with a Headset
2-5 times a week	once a week	2-5 times a week	very rarely	very rarely
very rarely	very rarely	very rarely	very rarely	very rarely
once a week	once a month	once a week	2-5 times a week	once a month
2-5 times a week	once a week	very rarely	very rarely	very rarely
daily	very rarely	very rarely	2-5 times a week	2-5 times a week
daily	daily	very rarely	very rarely	daily
very rarely	very rarely	very rarely	2-5 times a week	2-5 times a week
daily	2-5 times a week	very rarely	very rarely	very rarely
daily	2-5 times a week	once a month	very rarely	very rarely
once a month	2-5 times a week	daily	very rarely	daily
2-5 times a week	daily	once a month	very rarely	very rarely
very rarely	very rarely	very rarely	very rarely	very rarely
daily	2-5 times a week	very rarely	very rarely	very rarely
daily	once a week	very rarely	very rarely	very rarely
daily	once a week	2-5 times a week	2-5 times a week	2-5 times a week
daily	very rarely	daily	very rarely	very rarely
*Total	62.50%	37.50%	25%	31.25%

Legend

Very Rarely
Once a Month
Once a Week
2-5 Times a Week
Daily

Increasing Frequency

* We make an assumption that at least 2-5 times a week can be considered as frequent

Experience with Immersive Technology Platforms	Experience with IoT/Controlling Smart Devices
Headset (for research & games) phone (Snapchat, Pokemon Go, etc) not really	Controlling a TV with phone, usually to share videos with family. Google home, usually used to control music and lights- sometimes asking trivia questions to and/or weather. Phone camera (air drop), sharing photos with others in the room. nest: setting temperature of the home
AR- headset (HoloLens), snapchat filters	I have little to no experience with IoT devices and technologies and do not use IoT technologies on a regular basis.
snapchat filters, Samsung Gear VR, PSVR	I have a smart TV at home and a Google Home mini speaker. The speaker is new and I haven't had the chance to use it, but it will be used for music. I control the smart TV with my phone.
I have experience with VR, MR, AR on the HTC Vive, Windows mixed reality headset, HoloLens, Meta, Pixel 2 (android AR) Ipad Pro.	I have used: Google home mini but not often. I have used my phone to connect with a TV.
VR-headset. AR-phone	Lizard Smart Switch -> Controls lizards water and lightening from app. Google home mini -> set alarms, listen to news, control lizard switch
headset	I used to develop simple applications with HoloLens, Kinect.
Nope :)	TV cleaners Drones. TV and cleaners: for daily uses. Drones: for playing and camera capturing, etc.
Augmented Reality- Tablet and Phone	Chromecast- cast video (YouTube). Magic Mirror powered by a Raspberry Pi Displaying Weather for a set location (Manually Set)
VR: Games, Development, Oculus, Vive, mixed reality headset. AR: 305, Snapchat, HoloLens, iPad	Computer from computer, computer from remote, computer from phone, Google home, Amazon Alexa, Lights from phone
Used mobile apps for AR. VR rarely used. Used video game AR devices- Kinect.	Developed smart home energy management system- using Zigbee, circuit boards, wireless communication between monitoring devices and electrical devices.
used a snapchat filter once, but that's about it. I have used an AR device once in the past (30 minutes or so about 2 years ago).	While I do not own any such devices, I have encountered others using them, but have not used them myself.
I have tested/played with VR for MRI.	I have used my iPhone to control my Apple TV.
No	I have connected my phone to the TV once or twice.
iPad, HoloLens, HTC Vive	Controlling TV with iPad. Just for settings. Control cameras and the heating system/furnace.
I have had experience with virtual reality (from a friend's head set); it was game-based, e.g. throwing darts, catching objects, etc.	I have not had experience with IoT.

Appendix D: Raw Data

Legend

Gesture Mapping

Actions

Air Hold (as if holding something that is not seen (with space between fingers))

Air Tap

Bloom

Bring Together

Drag

Pinch

Release

Rotate

Spread Apart

Swipe (could also be like a push if it is using the whole arm)

Tap

Tap and Hold

Palm

A = Open Palm

A-L = Open palm facing left

A-R = Open palm facing right

A-F = Open palm facing forward

A-D = Open palm facing down

A-I = Open palm facing inwards towards the participant themselves

F = Fist

F4 = All four fingers

C-U = Closed palm facing up (may use some fingers as indicated by other letters)

Side

[L] = Left arm/hand

[R] = Right arm/hand

Fingers

T = Thumb

P = Pointer Finger

M = Middle Finger

RI = Ring Finger

K = Pinky

Supplements

->R = Action going from left to right

->L = Action going from right to left

X2 = Double tap (or other action).

Could be any number

D = Down

U = Up

V = Gesture supplemented with voice command

ARM = uses entire arm

C = Used paper towel before using gesture or voice command. *Only applies to cooking one.

Round 1: Connect Computer to Print Queue

Gesture

Participant	Next Options	Connect as Input	Pick Print Queue	Pick Printer 3	Collapse the Menu
1	air tap T, P [L]	air tap T, P [L]	air tap T, P [L]	air tap T, P [L]	swipe->R, A-R [L] ARM
3	tap P [R]	tap P [R]	tap P [R]	tap P [R]	tap P [R]
5	air tap T, P [R]	air tap T, P [R]	air tap T, P [R]	air tap T, P [R]	air tap T, P, X2 [R]
7	air tap T, P [R]	air tap T, P [R]	air tap T, P [R]	air tap T, P [R]	bloom [R]
9	tap P, X2 [L]	tap P, X2 [L]	tap P, X2 [R]	tap P, X2 [R]	tap P, X2 [R]
11	spread apart A-F [R]	tap P [R]	tap P [R]	tap P [R]	bring together A-F [R]
13	tap P [R]	swipe->U, F [R] ARM	tap P [R]	tap P [R]	swipe->D, A-D [R] ARM
15	tap P [R]	tap P [R]	tap P [R]	tap P [R]	tap P [R]

Voice

Participant	Next options	connect as input	pick print queue	pick printer 3	collapse the menu
2	"hit the button"	"connect as input"	"connect as input print queue"	"printer 3"	"collapse the print queue"
4	"right click the button"	"connect as input"	"print queue"	"printer 3"	"collapse the far right menu" ("or does it all just collapse?")
6	"open menu"	"select connect as input"	"select print queue"	"select printer 3"	"collapse all menus"
8	"click button"	"connect as input"	"print queue"	"printer 3"	"erase"
10	"click"	"click" (looking at button)	"click" (looking at button)	"select printer 3"	"click" (looking at initial button)
12	"printer open"	"connect as input"	"print queue"	"printer 3"	"collapse"
14	"click button"	"connect as input"	"print queue"	"printer 3"	"home"
16	"show options"	"connect as input"	"print queue"	"printer 3"	"collapse menu"

Round 1: Blinds Control

Gesture

Participant	Rotate to the Left	Rotate to the Right	Make It So That They Are Not Covering Window
1	tap A-F [R], air tap T, P [R]	air tap T, P [R]	swipe A-F, U [R] ARM
3	rotate->L C-U, P [R], pinch rotate->L C-U [R]	pinch rotate->R C-U [R]	tap and hold P [R]
5	air tap T, P, X3 [R]	air tap T, P, X4 [R]	swipe->R A-F [R] and swipe->L A-F [L] ARM
7	air tap T, P, X3 [R]	air tap T, P, X2 [R]	air tap T, P [R]
9	air hold and rotate->L T, P [L]	air hold and rotate->L T, P [L]	pinch T, P and swipe->D, T, P [L] ARM
11	tap T, P, X5 [R], tap and hold A-D [R]	tap and hold A-D [R], tap T, P, X2 [R]	swipe->R A-F [R] and swipe->L A-F [L] ARM
13	air hold and rotate->L T, F4, X10 [R]	air hold and rotate->R T, F4, X7 [R]	swipe->L A-L, [R] ARM
15	tap P, X3 [R], pinch and rotate->L T, P [R]	tap P, X3 [R], pinch and rotate->R T, P [R]	pinch T, P and swipe->R A-F [R] and pinch T, P and swipe->L A-F [L] ARM

Voice

Participant	rotate to the left	rotate to the right	make it so that they are not covering window
2	"move to left"	"move right"	"open them to right"
4	"rotate to the left using the button"	"rotate to the right using the button"	"open the blinds using the button"
6	"rotate blinds to the left"	"rotate blinds to the right"	"open blinds 100%"
8	"rotate the blinds to the left"	"rotate the blinds to the right"	"open all the blinds and move them all away" (+ also accompanied with both hands open in front moving away from each other)
10	"click"	"click"	"open"
12	"spin left"	"spin right"	"open window" (+also accompanied by parting hands with palms open)
14	"blinds rotate to the left"	"blinds rotate to the right"	"blinds open"
16	"rotate blinds to left"	"rotate blinds to right"	"open blinds"

Round 1: Speaker Control

Gesture

Participant	Pick the Song	Pause the Song	Play the Song
1	air tap T, P [R]	air tap T, P [R]	air tap T, P [R]
3	tap P [R]	tap P [R]	tap P [R]
5	air tap T, P [R]	air tap T, P [R]	air tap T, P [R]
7	air tap T, P [R]	air tap T, P [R]	air tap T, P [R]
9	tap P [L]	tap P [L]	tap P [L]
11	tap P [R]	swipe->D A-D [R] ARM	swipe->U A-D [R] ARM
13	tap P [R]	tap P [R]	tap P [R]
15	air tap T, P [R]	air tap T, P [R]	air tap T, P [R]/ tap P [R]

Voice

Participant	pick the song	pause the song	play the song
2	"play everywhere"	"pause"	"play"
4	"speaker please play song 1"	"pause the song"	"play song one"
6	"play song 1"	"pause song"	"play song"
8	"play song 1"	"pause"	"play"
10	"play song 1"	"pause"	"play"
12	"play song 1"	"pause"	"resume"
14	"play my favourite song"	"pause"	"play music"
16	"play song 1"	"pause song"	"play"

Round 1: Following a Recipe (Cooking)

Gesture

Participant	After Adding Water	After Adding Yeast	After Adding Sugar	After Adding Salt	After Adding Oil	After Adding Flour (and Kneading)	After Adding Water as Required	Kneads With?
1	air tap T, P [R]	air tap T, P [R]	air tap T, P [R]	air tap T, P [R]	air tap T, P [R]	C, air tap T, P [R]	air tap T, P [R]	both
3	air tap T, P [R]	tap P [R]	tap P [R]	tap P [R]	tap P [R]	rotate then tap P[R]	tap P [R]	both
5	air tap T, P [R]	air tap T, P [R]	air tap T, P [R]	air tap T, P [R]	air tap T, P [R]	air tap T, P [R] (did not knead yet)	(kneaded now) air tap T, P [L]	right
7	air tap T, P, X2 [R]	air tap T, P [R]	air tap T, P [R]	air tap T, P [R]	air tap T, P [R]	air tap T, P [R] (did not knead yet)	(kneaded now) air tap T, P [L]	right
9	tap P [R]	tap P [R]	tap P [R]	tap P [R]	tap P [R]	tap P [L]	tap P [L]	right
11	flip A-D-U [R]	flip A-D-U [R]	flip A-D-U [R]	flip A-D-U [R]	flip A-D-U [R]	flip A-U-D [L] ARM (left arm moved to right side of body)	flip A-U-D [L] ARM (left arm moved to right side of body)	right
13	tap P [R], V	tap A-F [R], V	tap A-F [R]	tap A-F [R]	swipe->L A-L [R] ARM	swipe->L A-L [R] ARM	swipe->L A-L [R] ARM	right
15	tap P [R], V	tap P [R]	tap P [R]	tap P [R], V	tap P [R]	tap P [R], V (did not knead yet)	(kneaded now) tap P [L]	right

**Note: Flip was also a kind of rotate.*

Voice

Participant	After adding water	After adding yeast	After adding sugar	After adding salt	After adding oil	After added dough and knead	After adding water as required	Kneads with?
2	ok "next"	"next"	ok "next"	"ok next"	"k next"	"next"	"next"	both
4	"next"	"next"	"next"	"next"	"k next"	"k next"	cleans hands with paper towel to add more water.	both
6	"next step" (also said "ok" before but was not meant for the system)	"next step"	"ok next step"	"k next step"	"k next step"	knead with right hand "next step"	"yeah next step"	right
8	"ok next"	"next"	"ok" big pause "next" maybe ok was meant to let the system know to go next	"next"	"next"	kneaded with right hand finger (actually just kind of stirred)"ok" pause and cleans hand with paper towel "next"	mixed with right hand pointer finger stir, cleans hand with paper towel, "next"	right
10	"k next step"	"next step"	"next step"	"next step"	"next step"	"next step" -"last step go back"	"next step"	both
12	"next step"	"next step"	"next step"	"next step"	"next step"	"next step"	"next step"	right
14	"next"	"next"	"next"	"next"	"next"	"next"	"next"	both
16	"uh next"	"uh next"	"uh next"	"uh next"	"uh next"	"ok next"	"ok next"	right

Round 2: Setting a Lights Schedule

Voice

Participant	next options	pick time	pick start	set hour to 7	set minutes to 25	set seconds to 30	set to AM	set as done
1	"hello computer...please program light time"	"hey Tim set time"	"start time tim"	"hours 7"	"minutes 25"	"and seconds to 30"	"AM"	"done"
3	"select"	"time"	"start"	"hours 7"	"minutes 25"	"seconds 30"	"AM"	"done"
5	"show more options"	"time"	"start"	"7 hours"	"25 minutes"	"30 seconds"	"AM"	"done"
7	"next"	"time"	"start"	"hours 7"	"25"	"30"	"AM"	"done"
9	"hey siri set a timer"	"time"	"start"	"set time for 7hours"	"25 minutes"	"and 30 seconds"	"AM"	"that's it"
11	"next"	"time"	"start"	"7 hours"	"25 minutes"	"30 seconds"	"AM"	"done"
13	"set the alarm (timer)"	"set the alarm (timer)"	"start"	"7"	"25"	"30 seconds"	"IN THE MORNING"	"TURN OFF"
15	"settings"	"set time"	"start time for"	"7"	"25"	"and 30 seconds"	"and AM"	"done"

Participant	pick end	set hour to 4	set minutes to 15	set seconds to 0	set to PM	pick day	set Monday	set Saturday	set as done
1	"end"	"hours to 4"	"minutes to 15"	"seconds are good"	"PM"	"day tim"	"lets set Monday"	"and Saturday"	"done"
3	"end"	"hours 4"	"minutes 15"	"seconds 0"	"PM"	"done" automatic open to day	"Monday"	"Saturday"	"done"
5	"end"	"4 hours"	"15 minutes"	"0 seconds"	"PM"	"done" "day"	"Monday"	"Saturday"	"done"
7	"end"	"4"	"15"	"0"	"PM"	"done" "day"	"Monday"	"Saturday"	"done"
9	"end time"	"4 hours"	"15 minutes"	"and 0 seconds"	"PM"	"and we're done" "day"	"Monday"	"Saturday"	"that's it"
11	"end"	"4 hours"	"15 minutes"	"0 seconds"	"PM"	"done" "day"	"Monday"	"Saturday"	"done"
13	"TURN OFF AT: "	"16:15: 30" SECONDS	(previous applies)	(previous applies)	TURN OFF AT 8:25:35 SECONDS. "IN THE AFTERNOON"	"TURN IT OFF"	"today" instead of "Monday"	"and Saturday"	"go to time"
15	"set up end time"	"hour to 4"	"15 minutes"	"0 seconds"	"4 hour 15 minutes PM"	"done" "set up day"	"Monday"	"and Saturday"	"done"

Gesture

Participant	Next Options	Pick Time	Pick Start	Set Hour to 7	Set Minutes to 25	Set Seconds to 30	Set to AM	Set as Done
2	swipe->L A-L [R] ARM	tap A-F [R] ARM	tap A-F [R] ARM	drag and tap P [R]	drag and tap P [R]	drag and tap P [R]	swipe->D A-F [R] ARM	swipe->D A-F [R] ARM
4	spread apart T, P [R]	spread apart T, P [R] and air tap T, P [R]	spread apart T, P [R]	pinch and drag->R, T, P [R]	spread apart T, P [R] (more like a click in place reverse air tap)	pinch and drag->R, T, P [R]	spread apart T, P [R]	air tap T, P [R]
6	tap P [R]	tap P [R]	tap P [R]	drag->R, P [R]	drag->R, P [R]	drag->R, P [R]	tap P [R]	tap P [R]
8	tap P [R]	tap P [R], V	tap P [R], V	tap P [R], V	tap and drag->R and tap P [R], V	tap and drag->R and tap P [R], V	tap P [R], V	tap P [R], V
10	swipe->R, T, M (ARM)	air tap T, P [R]	air tap T, P [R]	pinch and drag->R, T, P [R]	pinch and drag->R, T, P [R]	pinch and drag->R, T, P [R]	air tap T, P [R]	air tap T, P [R]
12	tap P [R]	tap P [R]	tap P, M [R] (used middle finger as well since this was the second option in the menu)	air hold A-I and tap P and drag->R [R] (used pointer finger as top for first slider)	air hold A-I and tap P, M and drag->R [R] (used middle finger as middle for second slider)	air hold A-I and tap P, M, RI and drag [R] (used ring finger as bottom for third slider)	tap P, M [R]	tap K [R]
14	circle with pointer finger *	tap P [R]	tap P [R]	tap P [R]	drag->R, P [R]	drag->R, P [R]	tap P [R]	tap P [R]
16	tap P [R]	tap P [R]	tap P [R]	drag->R, T [R]	drag->R, T [R]	drag->R, T [R]	tap P [R]	tap P [R]

Participant	Pick End	Set Hour to 4	Set Minutes to 15	Set Seconds to 0	Set to PM	Pick Day	Set Monday	Set Saturday	Set as Done
2	swipe->D A-F [R] ARM	air hold and drag->R and tap A-F T, F4 [R]	air hold and drag->R and tap A-F T, F4 [R]	air hold A-F T, F4 [R]	air hold and tap->D A-F T, F4 [R] ARM	tap A-F [R] ARM	tap A-F [R] ARM	tap A-F [R] ARM	air hold and tap->F A-F T, F4 [R] ARM
4	air tap T, P [R]	air hold and drag T, P [R] and air tap T, P [R]	air hold and drag T, P [R] and air tap T, P [R]	air tap T, P [R]	air tap T, P [R]	air tap T, P [R]	air tap T, P [R]	air tap T, P [R]	air tap T, P [R]
6	tap P [R]	drag->R, P [R]	drag->R, P [R]	tap P [R] (held in place)	tap P [R]	tap P [R]	tap P [R]	tap P [R]	tap P [R]
8	tap P [R]	drag->R P [R]	drag->R P [R]	tap P [R]	tap P [R]	tap P [R]	tap P [R]	tap P [R]	tap P [R]
10	air tap T, P [R]	pinch and drag->R, T, P [R]	pinch and drag->R, T, P [R]	no gesture. Just said "k"	air tap T, P [R]	air tap T, P [R]	air tap T, P [R]	air tap T, P [R]	air tap T, P [R]
12	tap P, M [R] (used middle finger as well since this was the second option in the menu)	air hold A-I and tap P and drag->R [R] (used pointer finger as top for first slider)	air hold A-I and tap P, M and drag->R [R] (used middle finger as middle for second slider)	air hold A-I and tap P, M, RI [R] (used ring finger as bottom for third slider)	air hold and tap->D A-F T, P [R]	air hold A-I and tap K [R] (used pinky finger for fun)	air hold A-I and tap K [R] (used pinky finger for fun)	air hold A-I and tap RI [R] (used ring finger for bottom)	air hold A-I and tap RI [R] (used ring finger for bottom)
14	tap P [R]	drag->R, P [R]	drag->R, P [R]	tap P [R]	tap P [R]	tap P [R]	tap P [R]	tap P [R]	tap P [R]
16	tap P [R]	drag->R, T [R]	drag->R, T [R]	tap P [R]	tap P [R]	tap P [R]	tap P [R]	tap P [R]	tap P [R]

Round 2: Thermostat Control

Voice

Participant	change to 24	change to 18	close
1	"Hey tim I'm cold can we change the hold at temperature to 24?"	"hey tim actually I'm way too hot now set it to 18"	"hey tim close out of that menu"
3	"select hold at 24"	"select hold at 18"	"exit"
5	"hold at 24 degrees celsius"	"hold at 18 degrees celsius"	"that's it"
7	repeatedly saying "plus"	repeatedly saying "minus" and at one point just said "minus 3"	"close it"
9	"hey Siri change the hold at temperature to 24"	"hey siri change the hold at to 18"	"hey Siri close the menu"
11	"change hold at 19 to 24"	"change hold at 24 to 18"	"close the thermostat control"
13	"set the temperature to 24"	"set the temperature to 18"	it should understand that it should close based on silence
15	"set up hold at 24 celsius"	"set up temperature to 18"	"k done close it"

Gesture

Participant	Change to 24	Change to 18	Close
2	tap A-F, X5 [R] ARM	tap A-F, X6 [R] ARM	tap A-F [R] ARM (reached to opposite left side)
4	air tap T, P, X5 [R]	air tap T, P, X6 [R]	air tap T, P [R]
6	tap P, X5 [R]/ air hold and rotate->R A-F T, F4 [R]	tap P, X6 [R]	tap P [R]
8	tap P, X5 [R]	tap P, X6 [R]	tap P [R]
10	tap P, X5 [R]	tap P, X6 [R]	tap A-F [R] ARM
12	air hold A-I and tap P, M, RI, X5 [R]	air hold A-I and tap K, X3 and air hold A-I and tap P, M, RI, X3P, [R]	tap A-F [L] ARM
14	tap P, X6 [R]	tap P, X6 [R]	tap P [R]
16	tap and hold P [R]	tap and hold P [R]	tap P [R]

Round 2: Video Display and Control

Voice

Participant	display 1	pick the video so it will play	close the video player so it will stop playing
1	"hey tim play video 1 on display 1"	"play it"	"Hey Tim stop the video"
3	"play video 1 on display 1"	"play video 1"	"pause. Stop" would pause before closing but nothing stopping you from directly closing
5	"place video 1 on display 1"	"play video on display 1"	"exit video player"
7	"go to screen 1"	"play"	"close it"
9	"hey siri play video 1 on display 1"	"hey siri play video on display 1"	"hey siri pause the video" (when asked again to stop instead of pause;) "hey siri stop mirroring to display 1"
11	"play video 1 on display 1"	"play video 1"	"exit from video 1"
13	"play this video on display 1"	it should automatically play from first prompt	"pause" "close the video" "shut it off" it should work will all of these. One day I want to say this other day I want to say that and it should always work
15	"play on display 1"	it should automatically play from first prompt. Or else "set up this video to display 1" and "play it"	"stop playing"

Gesture

Participant	Display 1	Pick the Video So It Will Play	Close the Video Player So It Will Stop Playing
2	swipe->L A-L [R] ARM	tap A-F [R]	tap P [R]
4	tap and drag->L and release T, P [R]	air tap T, P [R]	air tap T, P [R]
6	tap and drag->L, P [R]	tap P [R]	tap P, X2 [R] (pause and exit)
8	tap and drag->L and release P [R]	air tap T, P [R]	tap P [R]
10	swipe->L A-L [R] ARM	tap P [R]	tap P, X2 [R] (pause and exit)
12	swipe->L A-L [R] ARM	air hold A-I and tap P, M [R]	air hold A-I and tap P, M [R]
14	air hold and drag->L and release A-F, T, F4 [R]	tap P [R]	tap P [R]
16	air hold A-F, T, F4, and drag->L P [R]	tap P [R]	tap P [R]

Round 2: Fixing a Boombox

Voice

Participant	After Turn Boombox	After Pick Screwdriver	After Fixed Screwbit	After Place Screwbit	After Place Antenna	After Screw Antenna	After Lift Handle	After Tuck Antenna
1	"k tim go to the next step"	"next step tim"	"next step"	"next step tim"	"next step tim"	"next step tim"	"next step tim"	"next"
3	"uh next"	"next"	"next"	"next"	"next"	"next"	"next"	"next"
5	"k next step"	"next step"	"next step"	"next step"	"next step"	"next step"	"next step"	"next step"
7	"next"	"next"	"next"	"next"	"next"	"next"	"next"	"next"
9	"hey google I wanna fix this boombox" "whats the next step"	"ok next step whats next"	"whats the next step"	"whats next"	"whats next"	"ok whats next"	"done that whats next"	"whats next"
11	"next"	"next"	"next"	"next"	"next"	"next"	"back" "next"	"next"
13	"go to the next one"	"go to the next one"	"go to the next one"	"go to the next one"	"next"	"next"	"next"	"next"
15	"next"	"next"	"k next"	"done next"	"next"	"ok next"	"yeah next"	"yeah done next"

Gesture

Participant	After Turn Boombox	After Pick Screwdriver	After Fixed Screwbit	After Place Screwbit	After Place Antenna	After Screw Antenna	After Lift Handle	After Tuck Antenna
2	tap->D A-F [R]	tap->D A-F [R]	tap->D A-F [R]	tap->D A-F [R]	tap->D A-F [R]	tap->D A-F [R]	tap->D A-F [R]	tap->D A-F [R]
4	air tap T, P [R]	air tap T, P [R]	air tap T, P [R]	air tap T, P [R]	air tap T, P [R]	air tap T, P [R]	air tap T, P [R]	air tap T, P [R]
6	tap P [R], V	tap P [R], V	tap P [R], V	tap P [R], V	tap P [R], V	tap P [R], V	tap P [R], V	tap P [R], V
8	tap P [R]	tap P [R], V	tap P [R], V	tap P [R], V	tap P [R], V	tap P [R], V	tap P [L], V (right hand occupied)	tap P [R], V
10	swipe->L A-L [R] ARM	swipe->L A-L [R] ARM	swipe->L A-L [R] ARM	swipe->L A-L [R] ARM	swipe->L A-L [R] ARM	swipe->L A-L [R] ARM	swipe->L A-L [R] ARM	swipe->L A-L [R] ARM
12	tap A-F, T, F4 [R]	tap P, M [R] (holding screwdriver)	tap P, M [R] (holding screwdriver)	tap M, R, K [R] (holding screwdriver)	tap P, M [R] (holding screwdriver)	tap A-F, T, F4 [R]	tap A-F, T, F4 [R]	tap A-F, T, F4 [R]
14	tap P [R]	tap P [R]	tap P [R]	tap P [R]	tap P [R]	tap P [R]	tap P [R]	tap P [R]
16	tap P [R]	tap P [R]	tap P [L] (right hand occupied)	tap P [R]	tap P [R]	tap P [R]	tap P [R]	tap P [R] (next), tap P [R] (previous), tap P [R] (next)

Referent	Generic Participant Number		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
Close the Control Panel (Environmental Control)	Close Thermostat Control	Proposal 1	Proposed Voice Commands								tap	air tap	tap	tap	tap	tap	tap	tap
	Participant's Set		N/A								tap	air tap	tap	tap	tap	tap	tap	tap
	Final Set										tap, air tap							
	Final Set By Number of People who ever proposed gesture										tap (7), air tap (1)							
	Gesture Proposed by most amount of People										tap (0.86 agreement rate)							

Media Control

Referent	Generic Participant Number		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
Select Media to Play (Media Control)	Select Song	Proposal 1	air tap	tap	air tap	air tap	tap	swipe	tap	air tap	Proposed Voice Commands							
	Select Video	Proposal 1	Proposed Voice Commands								tap	air tap	tap	air tap	tap	air hold & tap	tap	tap
	Participant's Set		air tap	tap	air tap	air tap	tap	swipe	tap	air tap	tap	air tap	tap	tap	tap	air hold, tap	tap	tap
	Final Set										air tap, tap, air hold							
	Final Set By Number of People who ever proposed gesture										tap (10), air tap (6), air hold (1)							
		Gesture Proposed by most amount of People									tap (0.60 agreement rate)							

Referent	Generic Participant Number		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16
Select Button / Modify Playing Status (Media Control)	Pause/ Play Song	Proposal 1	air tap	tap	air tap	air tap	tap	swipe	tap	air tap	Proposed Voice Commands							
		Proposal 2	air tap	tap	air tap	air tap	tap	swipe	tap	air tap, tap								
	Stop Video	Proposal 1	Proposed Voice Commands								tap	air tap	tap	tap	tap	air hold & tap	tap	tap
	Participant's Set		air tap	tap	air tap	air tap	tap	swipe	tap	air tap, tap	tap	air tap	tap	tap	tap	air hold, tap	tap	tap
	Final Set										air tap, tap, swipe, air hold							
		Final Set By Number of People who ever proposed gesture									tap (11), air tap (5), air hold (1), swipe (1)							
		Gesture Proposed by most amount of People									tap (0.67 agreement rate)							

Referent	Generic Participant Number		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16		
Select Physical Display for	Select Video Display	Proposal 1	Proposed Voice Commands									swipe	tap & drag & release	tap & drag	tap & drag & release	swipe	swipe	air hold & drag & release	air hold & drag	
	Participant's Set		N/A									swipe	tap, drag, release	tap, drag	tap, drag, release	swipe	swipe	air hold, drag, release	air hold, drag	
	Final Set											swipe, tap, drag, release, air hold								
	Final Set By Number of People who ever proposed gesture											drag (5), tap (3), swipe (3), release (3), air hold (2)								
	Gesture Proposed by most amount of People											drag (0.57 agreement rate)								

Following a Workflow

Referent	Generic Participant Number		P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	
Go to Next / Previous Step (Following a Workflow) <i>Note: Not enough usage of previous</i>	Cooking a Recipe	Proposal 1	air tap	air tap	air tap	air tap	tap	rotate	tap	tap	Proposed Voice Commands								
		Proposal 2	air tap	tap	air tap	air tap	tap	rotate	tap	tap									
		Proposal 3	air tap	tap	air tap	air tap	tap	rotate	tap	tap									
		Proposal 4	air tap	tap	air tap	air tap	tap	rotate	tap	tap									
		Proposal 5	air tap	tap	air tap	air tap	tap	rotate	swipe	tap									
		Proposal 6	air tap	rotate & tap	air tap	air tap	tap	rotate	swipe	tap									
		Proposal 7	air tap	tap	air tap	air tap	tap	rotate	swipe	tap									
	Fixing a Boombox	Proposal 1										tap	air tap	tap	tap	swipe	tap	tap	tap
		Proposal 2										tap	air tap	tap	tap	swipe	tap	tap	tap
		Proposal 3										tap	air tap	tap	tap	swipe	tap	tap	tap
		Proposal 4										tap	air tap	tap	tap	swipe	tap	tap	tap
		Proposal 5										tap	air tap	tap	tap	swipe	tap	tap	tap
		Proposal 6										tap	air tap	tap	tap	swipe	tap	tap	tap
		Proposal 7										tap	air tap	tap	tap	swipe	tap	tap	tap
	Proposal 8										tap	air tap	tap	tap	swipe	tap	tap	tap	
	Participant's Set		air tap	air tap, tap, rotate	air tap	air tap	tap	rotate	tap, swipe	tap	tap	air tap	tap	tap	swipe	tap	tap	tap	
	Final Set		air tap, tap, rotate, swipe																
	Final Set By Number of People who ever proposed gesture		tap (10), air tap (5), rotate (2), swipe (2)																
	Gesture Proposed by most amount of People		tap (0.60 agreement rate)																

Appendix F: Voice Command Analysis

*Arbitrarily, more than 25% of users have to propose something in the same proposal round for it to be considered (at least 3 out of 8 in this case)

*If there are uncommon factors amongst the emerging command proposals of one task, those factors are removed from the emerging command of that task

*Furthermore, if there are uncommon factors amongst the emerging command of the two tasks, those factors are removed from the final command

Interacting with Menu System

[illegible][illegible][illegible][illegible]

Referent	Generic Participant Number	P9	P10	P11	P12	P13	P14	P15	P16	Emerging Command	Number of Users	
Collapse Menu (Interacting with Menu System)	Collapse Computer Options	Proposal 1	"collapse the print queue"	"collapse the far right menu"	"collapse all menus"	"erase"	"click"	"collapse"	"home"	"collapse menu"	"Collapse"/"Erase"/"Click" + "Menu"	7 out of 8
	Final Command - based on all proposals (& Number of Users Who Ever Proposed It)									"Collapse"/Other Verb ("Click"/"Erase" + "Menu"	7 out of 8	
	Agreement Rate									0.86		

Environmental Control

Referent	Generic Participant Number	P9	P10	P11	P12	P13	P14	P15	P16	Emerging Command	Number of Users	
rotate blinds	Proposal 1	show to left	"go to the left using the button"	rotate blinds to the left	rotate the blinds to the left	"click"	turn left	"blinds rotate to the left"	rotate blinds to left	"rotate /other verb ("spin"/ "move") + "blinds" + "left"	6 out of 8	
	Proposal 2	show right	"go to the right using the button"	rotate blinds to the right	rotate the blinds to the right	"click"	turn right	"blinds rotate to the right"	rotate blinds to right	"rotate /other verb ("spin"/ "move") + "blinds" + "right"	7 out of 8	
Pick Directional Button / Specify Direction (Environmental Control)	change thermostat temperature	Proposal 1	"hey tm i'm cold can we change the hold at temperature to 24"	green hold at 24	"hold at 24 degrees celcius"	repeatedly saying "plus"	"hey tm change the hold at temperature to 24"	green hold at 19 to 24	set the temperature to 24	go hold at 24 celcius	"change"/ other verb ("set"/ "select") + words on UI for temperature value ("hold at") + "temperature" + "24"	6 out of 8
		Proposal 2	"hey tm actually i'm way too hot now" green to 18	green hold at 18	"hold at 18 degrees celcius"	repeatedly saying "minus" and at one point just said "minus 3"	"hey tm change the hold at to 18"	green hold at 24 to 18	set the temperature to 18	go up temperature to 18	"change"/ other verb ("set"/ "select") + words on UI for temperature value ("hold at") + "18"	7 out of 8
Final Command - based on all proposals (6 Number of Users Who Ever Proposed It)										"rotate"/ "change"/ other verb ("spin"/ "move"/ "set"/ "select") + object ("blinds"/ "UI words for temperature") + value ("left"/ "right"/ numeric value)	14 out of 16	
Assessment Rate										v0.87		

Referent	Generic Participant Number		P9	P10	P11	P12	P13	P14	P15	P16	Emerging Command	Number of Users
Open Blinds Entirely (Environmental Control)	Open Blinds	Proposal 1	"open them to right"	"open the blinds using the button"	"open blinds 100%"	"open all the blinds and move them all away"	"open"	"open window"	"blinds open"	"open blinds"	"open" + "blinds"	8 out of 8
	Final Command - based on all proposals (& Number of Users Who Ever Proposed It)										"open" + "blinds"	8 out of 8
	Agreement Rate										1	

Referent	Generic Participant Number		P1	P2	P3	P4	P5	P6	P7	P8	Emerging Command	Number of Users
Close the Control Panel (Environmental Control)	Close Thermostat Control	Proposal 1	"hey tim close out of that menu"	"exit"	"that's it"	"close it"	"hey siri close the menu"	"close the thermostat control"	it should understand that it should close based on silence	"k done close it"	"close" + "it" / "Menu"	5 out of 8
	Final Command - based on all proposals (& Number of Users Who Ever Proposed It)										"close" + "it" / "Menu"	5 out of 8
	Agreement Rate										0.57	

Media Control

Referent	Generic Participant Number		P9	P10	P11	P12	P13	P14	P15	P16	Emerging Command	Number of Users
	Select Song	Proposal 1	"play everywhere"	"speaker please play song 1"	"play song 1"	"play song 1"	"play song 1"	"play song 1"	"play my favourite song"	"play song 1"	"Play" + "Song 1"	8 out of 8
Select Media to Play (Media Control)	Generic Participant Number		P1	P2	P3	P4	P5	P6	P7	P8		
	Select Video	Proposal 1	"play"	"play video 1"	"play video on display 1"	"play"	"hey siri play video on display 1"	"play video 1"	it should automatically play from first prompt	"play"	"Play" + "Video 1"/"Video"/"it"	8 out of 8
	Final Command - based on all proposals (& Number of Users Who Ever Proposed It)										"Play" + Media File Name	15 out of 16
	Agreement Rate											0.93

Referent	Generic Participant Number		P9	P10	P11	P12	P13	P14	P15	P16	Emerging Command	Number of Users
Select Button / Modify Playing Status (Media Control)	Pause / Play Song	Proposal 1	"pause"	"pause the song"	"pause song"	"pause"	"pause"	"pause"	"pause"	"pause song"	"Pause" + "Song"	8 out of 8
		Proposal 2	"play"	"play song one"	"play song"	"play"	"play"	"resume"	"play music"	"play"	"Play"	7 out of 8
	Generic Participant Number		P1	P2	P3	P4	P5	P6	P7	P8		
	Stop Video	Proposal 1	"Hey Tim stop the video"	"pause. Stop"	"exit video player"	"close it"	"hey siri pause the video"	"exit from video 1"	"pause"/"close the video"	"shut it off"	"stop playing"	8 out of 8
Final Command - based on all proposals (& Number of Users Who Ever Proposed It)											"Play"/"Pause"/"Stop"	13 out of 16
Agreement Rate											0.8	

Referent	Generic Participant Number		P1	P2	P3	P4	P5	P6	P7	P8	Emerging Command	Number of Users
Select Physical Display for Media (Media Control)	Select Video Display	Proposal 1	"hey tim play video 1 on display 1"	"play video 1 on display 1"	"place video 1 on display 1"	"go to screen 1"	"hey siri play video 1 on display 1"	"play video 1 on display 1"	"play this video on display 1"	"play on display 1"	"Play" + Media File Name + "on" + Hardware Name	7 out of 8
	Final Command - based on all proposals (& Number of Users Who Ever Proposed It)										"Play" + Media File Name + "on" + Hardware Name	7 out of 8
	Agreement Rate										0.86	

Following a Workflow

Referent	Generic Participant Number	P9	P10	P11	P12	P13	P14	P15	P16	Emerging Command	Number of Users	
Go to Next / Previous Step (Following a Workflow) <i>Note: Not enough usage of previous</i>	Cooking a Recipe	Proposal 1	"ok next"	"next"	"ok next step"	"ok next"	"k next step"	"next step"	"next"	"uh next"	[ok] + "next" + "step"	5 out of 8
		Proposal 2	"next"	"next"	"next step"	"next"	"next step"	"next step"	"next"	"uh next"	"next" + "step"	8 out of 8
		Proposal 3	"ok next"	"next"	"ok next step"	"ok next"	"next step"	"next step"	"next"	"uh next"	[ok] + "next" + "step"	5 out of 8
		Proposal 4	"ok next"	"next"	"k next step"	"next"	"next step"	"next step"	"next"	"uh next"	"next" + "step"	8 out of 8
		Proposal 5	"k next"	"k next"	"k next step"	"next"	"next step"	"next step"	"next"	"uh next"	[ok] + "next" + "step"	5 out of 8
		Proposal 6	"next"	"k next"	"next step"	"ok next"	"next step"	"last step go back"	"next step"	"next"	[ok] + "next" + "step"	6 out of 8
		Proposal 7	"next"	"next"	"yeah next step"	"next"	"next step"	"next step"	"next"	"ok next"	"next" + "step"	8 out of 8
	Task's Emerging Command										"next" + "step"	
	Fixing a Boombox	Proposal 1	"k tim go to the next step"	"uh next"	"k next step"	"next"	"hey google i wanna fix this boombox"	"next"	"go to the next one"	"next"	"next" + "step"	8 out of 8
		Proposal 2	"next step tim"	"next"	"next step"	"next"	"ok next step whats next"	"next"	"go to the next one"	"next"	"next" + "step"	8 out of 8
		Proposal 3	"next step"	"next"	"next step"	"next"	"whats the next step"	"next"	"go to the next one"	"k next"	"next" + "step"	8 out of 8
		Proposal 4	"next step tim"	"next"	"next step"	"next"	"whats next"	"next"	"go to the next one"	"done next"	"next"	8 out of 8
		Proposal 5	"tim can you go back"	"next"	"next step"	"next"	"whats next"	"next"	"next"	"next"	"next"	8 out of 8
		Proposal 6	"next step tim"	"next"	"next step"	"next"	"ok whats next"	"next"	"next"	"ok next"	"next"	8 out of 8
		Proposal 7	"next step tim"	"next"	"next step"	"next"	"done that whats next"	"back"	"next"	"yeah next"	"next"	8 out of 8
		Proposal 8	"next"	"next"	"next step"	"next"	"whats next"	"next"	"next"	"yeah done next"	"next"	8 out of 8
Task's Emerging Command										"next"		
Final Command - based on all proposals (& Number of Users Who Ever Proposed It)										"next"	16 out of 16	
Agreement Rate										1		

NOTE: "ok" was not meant as a part of the command, but since it emerged amongst multiple participants, it was analyzed as well. This is further discussed in chapter 4

Appendix G: Non-Abstracted (Detailed) Results Table

Referent	Context	Gesture (# of users)	Most Popular Gesture (total unique # of users) agreement rate	Voice Command Template (Order of Components Does Not Matter) (# of users)	Combined Voice Command Template for Each Referent (Order of Components Does Not Matter) (total unique # of users) agreement rate
Interacting with Menu System					
Expand Menu	expand computer options	tap (4) air tap (3) spread apart (1)	tap (9 out of 16) 0.53	"click"/other verb ("hit"/"open") + "button" (7)	"click"/"set"/ other verb ("hit"/"open"/"select"/"program") + object ("button"/"timer") (11 out of 16) 0.67
	expand light options	tap (5) swipe (2) air tap (1)		"set"/other verb ("select"/"program") + "timer"/"time" (4)	
Pick Button	printer menu buttons	air tap (3) swipe (1)	tap (11 out of 16) 0.67	exact words on button (8)	exact words on button (16 out of 16) 1
	light menu buttons	tap (5) air hold (2) air tap (2)		exact words on button (8)	
Set Slider	set light time sliders	drag (8) tap (4) air hold (3) pinch (2) air tap (1) no gesture (1)	drag (8 out of 8) 1	slider name + value (8)	slider name + value (8 out of 8) 1
Set Toggle	toggle on light schedule days	tap (6) air hold (2) air tap (2) swipe (1)	tap (6 out of 8) 0.71	toggle value (8)	toggle value (8 out of 8) 1
Collapse Menu	collapse computer menu	tap (3) swipe (2) air tap (1) blow (1) bring together (1)	tap (3 out of 8) 0.29 *low agreement rate	"collapse"/ other verb ("erase"/"click") + "menu" (7)	"collapse"/ other verb ("erase"/"click") + "menu" (7 out of 8) 0.86
Environmental Control					
Pick Directional Button/ Specify Direction	rotate blinds	rotate (4) air tap (3) tap (3) pinch (2) air hold (2) tap and hold (1)	tap (9 out of 16) 0.53	"rotate"/other verb ("spin"/ "move") + "blinds" + "left"/ "right" (7)	"rotate"/"change"/ other verb ("spin"/"move"/"set"/"select") + object ("blinds"/ UI words for temperature) + value ("left"/"right"/numeric value) (14 out of 16) 0.87
	change hold at temperature	tap (6) air hold (2) air tap (1) rotate (1) tap and hold (1)		"change"/ other verb ("set"/ "select") + words on UI for temperature value ("hold at") + numeric value (7)	
Open Blinds Entirely	open blinds	swipe (6) pinch (2) air tap (1) tap and hold (1)	swipe (6 out of 8) 0.71	"open" + "blinds" (8)	"open" + "blinds" (8 out of 8) 1
Close the Control Panel	close thermostat control	tap (7) air tap (1)	tap (7 out of 8) 0.86	"close" + "menu"/"it" (5)	"close" + "menu"/"it" (5 out of 8) 0.57
Media Control					
Select Media to Play	play song	air tap (4) tap (4)	tap (10 out of 16) 0.6	"play" + "song 1" (8)	"play" + media file name (15 out of 16) 0.93
	play video	tap (6) air tap (2) air hold (1)		"play" + "video 1"/"video"/"it" (7)	
Select Button /Modify Playing Status	play/pause the Song	air tap (4) tap (4) swipe (1)	tap (11 out of 16) 0.67	"play"/"pause" (8)	"play"/"pause" / "stop" (13 out of 16) 0.80
	stop/pause the video	tap (7) air hold (1) air tap (1) drag (1)		pause"/"stop" + "video" (5)	
Select Physical Display for Media	play specific video on specific display	release (3) swipe (3) tap (3) air hold (2)	drag (5 out of 8) 0.57	"play" + media file name + "on" + display name (7)	"play" + media file name + "on" + display name (7 out of 8) 0.86
Following a Workflow					
Go to Next/ Previous Step (not enough users used previous)	go to next step while cooking	air tap (4) tap (4) rotate (2) swipe (1) tap (6)	tap (10 out of 16) 0.6	"next" + "step" (8)	"next" (16 out of 16) 1
	go to next step while fixing boombox	air tap (1) swipe (1)		"next" (8)	

Referent table with the resulting voice command set and gesture set. This table details two different contexts for some referents and the results of each of those. The results of the two contexts are combined to form a final result (for both voice and gesture). The number of users who made each proposal can be seen here. The agreement rates are seen in the final voice and final gesture columns.

Appendix H: Storyboards

The initial breakdown of all the storyboards. Some tasks were modified or removed later.

①

① Interacting with Menu System

- * Visualizations
- * Configure a schedule for lights (complex)
- * Printer example

② Follow Workflow (→ info access)

- * Carbon Monoxide (include AR keypad?)
- * Fix a toilet
- * Cook / Follow a recipe (ie kneading)

③ Media Control

- * view media {attach?}
- * Capture media →

④ Environmental Control (physical environment)

- * Controlling heating / lighting for building
 - ↳ color
 - ↳ on/off
- * Unlocking home? - blinds - location of menu.

Each scenario had 2 tasks (contexts), and each context was storyboarded 3 times with different variations. The sample storyboard below is the thermostat control task (context) storyboard 1.