In-Air Gestures in the Home Environment

Karen Ho
University of California,
San Diego, U.S.A.
kho742@gmail.com

Hanley Weng
University of Sydney,
Australia
hwen6932@uni.sydney.edu.au

ABSTRACT

The home serves as a place of safety, privacy and comfort. We interact with this space daily, and as a result, many gestures arise as we adjust our environment to our needs. In this paper, we discuss the implications of in-air gestures in the home environment as they relate to feedback response time. We test three iterations of prototypes with user studies. We conclude with our findings and where this research can grow in the future.

Author Keywords
in-air gestures; feedback delay; home; remote control; hand gestures.

ACM Classification Keywords
H.5.2. Information Interfaces and Presentation: User Interfaces - *Input devices and strategies, Interaction styles, Prototyping*. D.1.6. Programming Techniques: Programming Environments - *Interactive environments*.

INTRODUCTION
The home today serves as a place of safety and security. We interact with many of the fixtures on a daily basis to make it more comfortable for our needs. As a result, we have created many gestures that we associate with fixtures throughout the home. In this paper we will go through three iterations of user studies and prototyping techniques to discuss the gestures that users in a home environment prefer and how they relate to the behavior of the task. Our initial focus was on the home environment, so we began with a user survey on multi-modal home interactions. After this survey we narrowed our focus down to unimodal in-air gestures, and implemented a basic in-air gesture responsive prototype and conducted user studies of gesture type preferences and feedback delay preferences. With these findings, we created our third and final iteration, which incorporated more context and a refined gesture set. The user study of this leads us to our conclusions about designing gestures and possible directions for future work.

Home Environment

Many interactions already occur in the home, and as a result many systems have been made for systems incorporated into the home. Prior research into home control systems have stated the benefits of home automation systems and having remote interfaces to control the home. These systems are especially helpful for the elderly or mobility impaired, where physically accessing an interface is more difficult than intended. Home automation systems have taken several courses in their technological development. With the Kymera Wand [11], users can ideally control any fixture in their homes by gesturing with a physical "wand". Other systems that incorporate one single central screen for universal home control have also been explored, but they restrict the portability aspect of the previous system. Systems like the Gesture Pendant seen in "The Gesture Pendant: A Self-illuminating, Wearable, Infrared Computer Vision System for Home Automation Control and Medical Monitoring" by Thad Starner et al. can control the home remotely with hand gestures detected by a sensor worn by the user. This eliminates the need for a physical remote device as seen with the Kymera Wand, the user still needs to ensure that the sensor is being worn at all times in order for the system to work. Starner et al. also argue that "While speech recognition has long been viewed as the ultimate interface for home automation, there are many problems in this domain." [8], some of which include the problem with noise and the social aspects of speaking to one's house when others are around. The Gesture Pendant also has the advantage over a universal home control system in that the interactive points tend to be difficult to see or learn [8]. All of these systems have potential for technical growth, but each has its own downside.

Systems for the Handicapped at home

The three previous systems have benefits that can be found in environments used by people with handicaps or mobility-impairments. By having the option to remotely control fixtures throughout the home, users do not have to exert physical strain or difficulties to interact with systems through their original or intended interface. Another system fully intended for handicapped users is found in research in Brain Computer Interfaces (BCI). These are a possible interaction technique for people with complete physical handicaps, where any body movement gesture is not possible. The problem with this technique lies mostly with the technical capabilities of this system. Currently speed, reliability, and versatility are all lacking in comparison to other interaction techniques, but as technology grows, BCI can become a more viable solution for home control, especially for fully handicapped users [10].

RELATED WORK

In-air gestures in the home environment builds upon feedback in a physical space, gestural communication, gestures as a separate paradigm to touch, and what has been done to address the live-mic problem.

Feedback in the physical space

Changes to a physical environment through digital interfaces have often directly been manipulated through a single visual display. A display that does not always reflect a user's mental model of their environment. Augmented reality (AR) attempts to remedy this by overlaying an augmented view of information over regions of interest. However, many mobile AR systems rely on the user holding a portable device, and deter the free use of hands. This issue is being recognized and remedied by emerging AR systems attempt such as Sixth Sense [5] and near-retina display technologies such as Google Glasses [12] effectively projects this layer of information whilst keeping a user's hands free to interact with their environment. [6]

Communicating with gestures

Humans use gestures everyday as a form of nonverbal communication. Gestures can communicate joint attention, establish space, communicate emotion, and language. Gestures can even have specific symbolic meanings, from emotional expressions to sign language. There is much research in computer vision that attempt to decode sign-language to allow for digital communication for the deaf [4, 9]. Sean Gustafson, Daniel Bierwith and Patrick Baudisch have also shown that gestures can viably communicate with invisible "Imaginary Interfaces" (2010) [2]. There have also been recent technological leaps (e.g. Leap 3D Motion [13]) that allow for the recognition of very fine finger and hand detection.

In-Air Gestures are not Touch

The paradigm of three-dimensional in-air gestures cannot be simply treated as a direct translation of two-dimensional touch interactions. Much like how touch is a separate interaction paradigm to mouse-input, in-air gestures must be allowed to develop its own unique input/output paradigm. [1]

Unlike touch, systems that read in-air gestures are always on, and subject to the "live mic" problem. This is a prevalent problem that has been explored widely explored in voice-input systems. Del Ra and William [1] suggest three ways of dealing with this; the use of reserved actions (design gestures that are highly unlikely to be accidentally triggered by the user unless desired), reserved clutches (a special gesture to initiate input commands), multimodal input (e.g. other modalities such as buttons, voice, to initiate the system). Different systems have used different solutions, traditional versions of OSX (prior to version 10.8), could require the user to call out the name of the computer prior to any commands (a reserved clutch). Siri on the iPhone 4s requires the initial holding of a button to initiate the voice-command-system (multi-modal input). Other design projects such as Jaime Ruiz' Double Flip [7] used a reserved clutch, termed as a delimiter, to initiate gestural navigation.

In regards to "Touch versus in-air Hand Gestures", A. Hassani notes that, though there is generally an equal preference for touch or in-air gestures amongst seniors, this is strongly dependent on context. Seniors also state a preference for in-air gestures in tasks related to remote control [3].

USER STUDIES

1. USER SURVEY: MULTIMODAL HOME INTERACTIONS

Introduction

Our concept initially centered on multimodal home interactions. Our initial research into novel interaction techniques in the home showed that gestures have been explored in controlling specific fixtures in the house and using multimodal techniques would be a viable research point. The control of appliances and fixtures around the house allows for an increase in awareness, which touches on privacy and safety issues that are prominent in a home environment. The purpose of this initial user survey was to explore interactions in the home environment and how the functions of the "home" as an interface relates to how people interact with their surroundings.

Tasks

The task observed in this study was to discover user-defined gestures in the appropriate context. These surveys were conducted in the participant's own homes. We began by priming our participants with a context akin with imagining everything in their home environment as remotely controllable by gestures, voice, and other forms of physical output that did not include readable brainwaves. We then asked closed questions like "how would you control <particular fixture>?" regarding various specific appliances and fixtures like lights, doors, televisions, and windows. We then opened up the discussion and asked more broad questions to acquire more insight into what areas people would prefer to use a remote gesture over the current existing interface.

Participants

In this survey we interviewed 10 participants (4 female), all of whom were college students of varying backgrounds and disciplines. Our findings showed that even though they were primed with a multimodal context, all of our participants preferred unimodal input.

Results

9 of the 10 participants preferred gesture-only over voice-only input. Most gestural preferences were localized in the arm and commonly mimicked skeuomorphism, miming the common existing physical interaction with the fixture or

incorporating an abstracted representation of the movement of the fixture (e.g. one user utilised a form of sign language to communicate their commands).

Discussion
Even in a multimodal environment, users preferred uni-modal input techniques (either voice-only or body gesture-only). Based on our initial user interviews, we decided to narrow our research even more and focus on in-air hand gestures.

2. USER STUDY: GESTURE MODE AND DELAY

Introduction

With our research narrowed to in-air gestures in the home environment, we decided to investigate the feedback aspect of interaction. The purpose of this user study was to find 1) if there was a preference over in-air gestures, as varied by execution time, physical effort, direction of motion, associated sound cues, and the number of hands involved in the gesture, and 2) whether users had a preference for feedback as varied by response time.

Tasks

The task consisted of two parts: 1. User satisfaction of gestures: In this part we created 30 gestures comprised of combinations of varied direction of motion/rotation (inward, outward, forward, back), physical effort (finger point, forearm push, full arm sweep), execution time (instantaneous snap, continuous sweep), and associated sound cues (snap vs point, clap vs wave). 2. User satisfaction of feedback delay: In this part we used one defined gesture (open-palm forward push) and randomly varied the feedback response delay from a range of 0 to 3 seconds.

System

The system was a simple input/output application that responds to a body moving beyond a distance threshold as detected by a Microsoft Kinect. The output was a change in monitor display screen color to indicate to the user that a body had moved past the specified threshold. For the survey, users were given an iPad to fill in a Google Docs Survey. Data from the system, the order of gestures and values of delays, were later matched with this digital survey.

Participants

We had 6 participants do both tasks, all of whom were right-handed college students of various backgrounds and disciplines.

Experiment

In the experiment we randomized the order between task 1 and 2. Task 1 had users try each of the 30 unique gestures we defined and was kept as an open discussion, where users could redo or retry any gesture at any time. Participants ranked each of gestures from 1 (least) - 7 (most) in satisfaction. In task 2, we conducted 4 sets of 4 random feedback delays ranging from 0-3 seconds and had our participants rank from 1 (least) to 7 (most) satisfaction.

Results

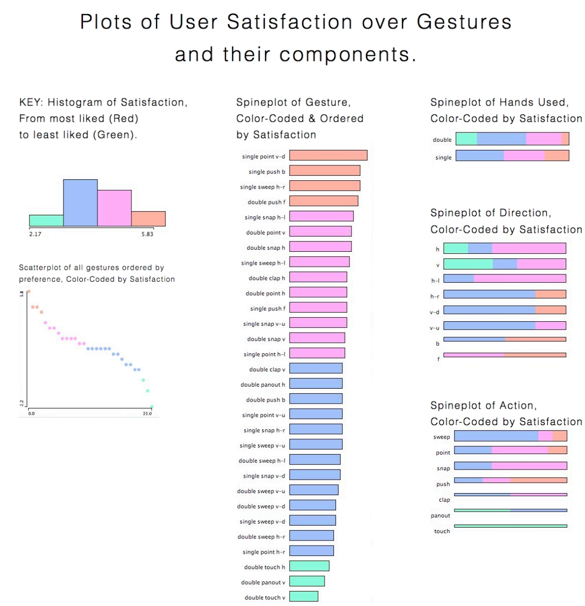

Figure 1. Plots of user satisfaction of gestures and their characteristics; hands-used, direction, action. User-Satisfaction is segmented into 4 parts for legibility; red being highest ranked in satisfaction, and green being the lowest. Terms: h = horizontal, v = vertical, -l = to the left for a right-handed user, -r = to the right for a right-handed user, d = down, u = up, b = backwards, f = forwards.

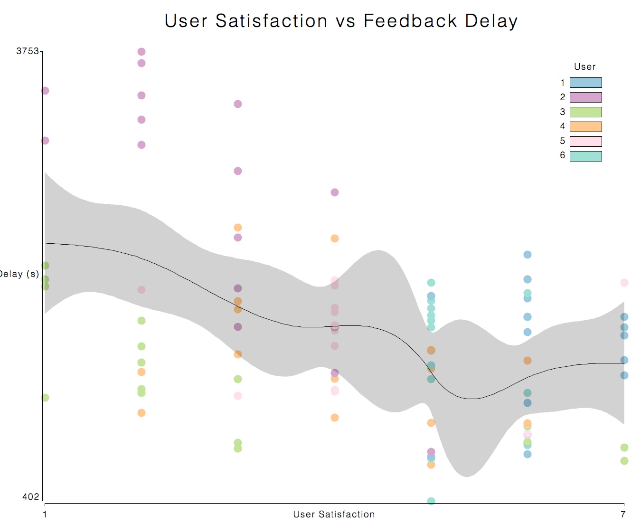

Figure 2. A scatterplot of the user satisfaction of the six users across feedback delay is represented. A spline (of degree; number of user satisfaction categories - 1) is shown. There is a significant increase in user satisfaction at time (1545 ± 188 ms). It should be noted that user 2 had a range of random delays from 0 to 4 seconds.

Discussion

From the results of Task 1, we found that our users preferred smaller gestures that involved less physical effort. This entails 1-handed gestures over 2-handed, motion centered around the wrist or fingers as opposed to the elbow or shoulder, and quick execution time.

From Task 2, we observed higher user satisfaction (5-7) at shorter delay times. The threshold where user satisfaction has a noticeable switch is around 1.5 seconds. This likely means that users begin noticing or feel dissatisfied with feedback delays longer than 1 second.

3. USER STUDY: FEEDBACK TRANSITIONS

Introduction

The purpose of this user study was to find how user preference for different transitional feedback times differed with gestures of different durations. Participants were tasked with discovering their preferred lighting fade-in times for a lamp, under an instantaneous 'shooting' gesture and a longer 'sweep gesture'.

Tasks

Participants were sat down in front of our apparatus, a customized lamp, and asked to perform a specific gesture to switch the lamp on and off. A tablet-device was used to present participants with a digital slider that they used to manipulate the fade-in duration of the lamp. Once the user had adjusted the slider to their ideal fade duration, this process was repeated with a second gesture. The two gestures presented to participants to emulate were an instantaneous 'point-and-shoot' gesture, and a longer 'sweep' gesture. The ordering of gestures alternated per user. Participants were then asked about their preferred gesture, and on their preferred fade-durations in comparison to faster and slower durations. The design of an ideal gesture was also asked of the participant, along with its corresponding fade-duration.

System

Our system consists of a lamp with three LEDs connected to an Arduino, which is connected to a laptop, running a Java program - which acts as the central-hub of the system. An iPad running the app TouchOSC was used to create a simple slider interface that sends messages to the laptop via wifi whenever the interface is interacted with. The interface on the iPad also had a 'record' button for the experimenters, which would signal the central java program on the laptop to record all incoming messages (i.e. all user manipulations with the slider). The slider adjusted a fade-duration variable between zero to five seconds. A Kinect was also plugged into the laptop and communicated directly with the main program, determining user-gestures. When a user-gesture was detected, the main program would tell the Arduino to adjust the brightness of all three LEDs within the lamp. The gesture-recogniser was based on hand-movement detection. The hand was detected by identifying a user, followed by the detection of an outreached hand (from the depth-histogram of the user). Face tracking was also added to ensure the user was looking at the system. A wizard-of-oz system was also implemented for fade-in-control for environments that did not have sufficient lighting for the system to function at full capacity.

Participants

5 participants (2 female) were recruited at our institution to take part in the study. They were aged 19 to 25. Participants were given a small gratuity for their time.

Results

user \ gesture	wave fade (s)	shoot fade (s)	average fade (s)	preferred
1	1.3798219	0.9718102	1.17581605	wave
2	0.21513343	0.21513343	0.21513343	shoot
3	1.4985162	0.296735765	0.897625983	wave
4	0.15243888	0.1905489	0.17149389	wave
5	1.46341475	0.82317085	1.1432928	wave
average	0.941865032	0.499479829	0.720672431	
standard deviation	0.693724246	0.369211391	0.493518744	

Figure 3. A table of user's preferred fade-in durations across two different gestures.

Most users preferred a wave gesture over shooting, and the shooting gesture had a typically short fade-duration preference in comparison to the wave gesture.

Discussion

After this study, as a general overview of our results, overall preference for feedback response time depends on the person. Notes taken within the study and post-study interviews revealed that factors such as mood, feelings and general personality that affected the temporal desires of the user all contributed to preference. It was also observed that users whom were less energetic during the study preferred longer duration times, this observation may correlate with the age of a user. Users did tend to match the fade duration similarly to user's hand movement duration. A desire of direct lamp-brightness-to-hand-position mapping was observed, especially in follow-up questions. Also, a transitional effect between light-modes (on/off) was preferred over instant (0.00s) fades. We postulate that this may correlate with the duration of hand gesture, or be simply a preferred feedback aesthetic.

IMPLICATIONS FOR DESIGN

Summarizing our results for user-interaction with fixtures in the home environment; we observed a preference for unimodal input even in a multimodal context. Users preferred gestures that required minimal physical effort, whilst maintaining confidence in the recognition system. preferred delay times varied greatly according to user and context. There was an apparent preference for topographic mappings between gestures and feedback. It was also perceived that gestures and their durations correlated with user-energy (i.e. including maturity and age).

Through our results and design process, we suggest three implications for in-air gestural design:

- **Natural initiations of gesture recognition.** Systems that are continuously on suffer from the 'live-mic' problem. One suggested solution is the use of a delimiter [7] to minimize false positives and conserve energy usage of the system. From observations of our participants, we propose delimiters that are natural, and near-invisible to the user. An example of a natural delimiter that is an extension of our implemented one, is the use of gaze-tracking and pointing to determine that the user is indeed looking at, and pointing at an object that they intend to command.

- **Topographically mapped gestures**. The personal mental models of the users should correspond with the feedback. Whether that feedback is continuous or discrete, inward or outward, the users' own mental models, and abstractions of it, best reflect the ideal characteristics of their gestures.

- **Context and demographic matters.** The key take away from our research was that context always matters. Depending on the situation and the task at hand, users will behave differently with interfaces in their environment. If they know the expected output, they will alter their input to allow the output to reflect any desired adjustments. When designing gestures for interfaces that people interact with regularly, we have to consider people's behaviors and understanding of systems in their environment.

CONCLUSION AND FUTURE WORK

Technologically, in future work, our system can look to advances in computer vision through neuroscience. Implementing systems of place fields, optical flow, synaptic plasticity, image compression, to better abstract movement in the system. Improving the accuracy of the system with our current hardware (a combination of depth sensors, infrared, and rgb-camera data, microphones), will allow us to generate accurate training data for less-advanced systems (such as a simple rgb-camera). The abstraction of the system to more basic systems, such as hardware that exists on common devices (e.g. a laptop's webcam and microphone) will allow us to more easily deploy the system. This deployment will allow the concept of in-air gestures to be tested on more systems as a technological probe, whilst evaluating and improving the system, through designs that naturally minimise false positives (e.g. natural delimiters).

In the realm of in-air gestures, our studies indicate that users tend to prefer smaller and faster hand gestures that require minimal physical exertion while maintaining a balance between user-system confidence and physical effort. Although we observed a correlation between feedback-delay of a gesture and user satisfaction, there is insufficient data to determine any true conclusions. Along with feedback-fade-in time, feedback-delay tests will benefit from additional data, under more specific contexts. In our research only one small aspect of interfaces in the home environment, feedback response time, was considered. There are many other aspects influencing how people may interact with a particular object such, one of the most prominent being the context of the fixture, its relative position, functionality, and characteristics in relation to the user. More expansive future work can explore the influences of multimodal feedback, in-air gestures outside of the home environment, in-air gestures through existing devices (e.g. phones), and gestural control over multiple dimensions and parameters.

Currently in-air gestures have been proposed as an alternative to everyday activities, and a benefit to those whom have their mobility disabled. However, such novel interactions have many yet-unexplored benefits, granting it much potential to enrich lives.

ACKNOWLEDGEMENTS

We would like to thank all our participants, and Dr. Anne Marie Piper for her input into this paper.

REFERENCES

1. Del Ra, III, William. 2011. Brave NUI world: designing natural user interfaces for touch and gesture by Daniel Wigdor and Dennis Wixon. *SIGSOFT Softw. Eng. Notes* 36, 6 (November 2011), 29-30.

2. Gustafson, Sean, Bierwirth, Daniel, and Baudisch, Patrick. 2010. Imaginary interfaces: spatial interaction with empty hands and without visual feedback. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology* (UIST '10). ACM, New York, NY, USA, 3-12.

3. Hassani, A.Z. (2011) Touch versus in-air Hand Gestures: Evaluating the acceptance by seniors of Human-Robot Interaction using Microsoft Kinect. Lecture Notes in Computer Science, 2011, Volume 7040/2011, 309-313.

4. Liang, Rung-Huei, Ouhyoung, Ming. "A real-time continuous gesture recognition system for sign language ,"*Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on* , vol., no., pp.558-567, 14-16 Apr 1998.

5. Mistry, Pranav and Maes, Pattie. 2009. SixthSense: a wearable gestural interface. In *ACM SIGGRAPH ASIA 2009 Sketches* (SIGGRAPH ASIA '09). ACM, New York, NY, USA, , Article 11 , 1 pages.

6. Purcher, Jack. Google Reveals Video Glasses Working with Magic Rings & Invisible Tattoos. 17 May, 2012**.** http://www.patentbolt.com/2012/05/google-reveals-video-glasses-working-with-magic-rings-invisible-tattoos.html.

7. Ruiz, Jaime and Li, Yang. 2011. DoubleFlip: a motion gesture delimiter for mobile interaction. In*Proceedings of the 2011 annual conference on Human factors in computing systems* (CHI '11). ACM, New York, NY, USA, 2717-2720.

8. Starner, T., Auxier, J., Ashbrook, D., Gandy, M., The gesture pendant: A self-illuminating, wearable, infrared computer vision system for home automation control and medical monitoring. *The Fourth International Symposium on Wearable Computers* (2000) 87-94.

9. Vogler, Christian; Metaxas, Dimitris. A Framework for Recognizing the Simultaneous Aspects of American Sign Language, Computer Vision and Image Understanding, Volume 81, Issue 3, March 2001, Pages 358-384, ISSN 1077-3142.

10. Wolpaw, J. R., McFarland, D.J., Vaughan T.M., Brain-Computer Interface Research at the Wadsworth Center. *IEEE Transactions on Rehabilitation Engineering,* Volume 8, Number 2, June (2000). 222-226

11. "The Wand Company - Brief Manual." 2009. 12 Jun. 2012 <http://www.thewandcompany.com/Manual.html>

12. "Google Reveals Video Glasses Working with Magic Rings - Patent Bolt." 2012. 12 Jun. 2012 <http://www.patentbolt.com/2012/05/google-reveals-video-glasses-working-with-magic-rings-invisible-tattoos.html>

13. "Leap Motion." 2012. 12 Jun. 2012 <http://www.leapmotion.com/>