EXPLORING AN INTELLIGENT RESPONSIVE ARCHITECTURE THROUGH GESTURE-BASED INTERACTION

by

Yijing An

A thesis submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Master of Architecture
and Master of Information Technology

Charlotte

2019

Approved by:

____
Dr. Dimitiros Papanikolaou

____
Dr. Celine Latulipe

____
Dr. Kaus Sarkar

ABSTRACT

YIJING AN. Exploring an Intelligent Responsive Architecture Through Gesture-based Interaction. (Under the direction of DR. DIMITRIOS PAPANIKOLAOU)

Architects are increasingly adopting gesture-based interaction to create responsive, engaging and inspiring spaces. However, there is no established theories providing guidelines to link gestures and spatial movement. This thesis critiques the prevailing approach in HBI in which users learn “cookie-cutter” gestures from gesture elicitation studies for interacting with the built environment. Instead, I argue that these gestures should emerge naturally as mutual convergence between users and intelligent architecture in order to be customized to different user group.

In this thesis, I demonstrate how gestures can emerge through the interaction between user and architecture by developing an interactive art installation with machine learning algorithm and confirm its potential of increasing curiosity and engagement by conducting a user study.

This thesis also compares these emerged gestures with the “cookie-cutter” gestures and raise the discussion on the implications for the future of intelligent responsive architecture.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

SVM	supportive vector machine
RQ	research question
HBI	human building interaction
OAA	one-against-all

CHAPTER 1: INTRODUCTION

"Advances in artificial intelligence would soon give rise to buildings capable of intelligently recognizing the activities of their users and responding to their needs."

---- Nicholas Negroponte, 1970

The term "responsive architecture" was introduced by Nicholas Negroponte during the late 1960s. This terminology was proposed when cybernetics was being applied in architecture design. Following Nicholas's contribution, each year, many new technologies appear to help architects gain access to human movement data and design responsive architecture to fulfill human needs. With the ability to record and observe this nonverbal, complex, rich body language, more and more architects are trying to study the relationship between human movements and space movement patterns. Interactive installations which can be operated through human inputs such as touch, gesture, voice, are developed for this purpose (Di Cristina, G., 2002).

Followed by these projects, one question that people barely ask but very important is: How should architecture respond to user's needs. It's easier to say: "I will make the wall move in response to human" than actually program it to move because it is hard to capture user's intention. How do you decide if the wall should move towards people or away from people? A simple example that fails to address this issue will be the automatic door. The door will always open when it senses the motion no matter the user wants to enter the door or just passes by. This interaction may embarrass the user and waste millions of dollars on air conditioning energy every year just because the question is not well addressed in the design phase.

Architects tried to address this question in two different approach.

First, architects try to pre-define the relationship. And whether easy or not, the user is asked to adapt. (Thiruvengada, H. et al, 2013).

For tangible interaction, the relationship between user's reaction and form of architecture has already been fully explored by Don Norman. His work provides the guidelines of how the form of architecture can be designed to fulfill human needs. (Norman, D., 2013) However, for intangible interaction, such as gesture-based interaction and voice-based interaction, there is no such established theories to link gesture or voice with the movement of architecture. Some architects choose to follow the existing rules in tangible

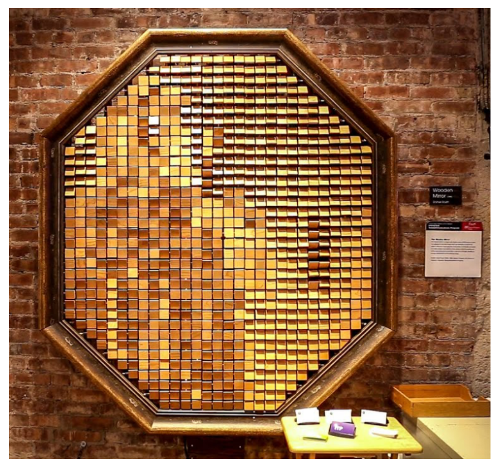


Figure 1: Cityhome - a Gesture-based home automation system (MIT's Media Lab)
Figure 2: Interactive wooden mirror (Daniel Rozin,1999)

interactions. In the CityHome project developed by MIT Media lab, user is taught to make "push" or "pull" gesture to open a drawer which is just like how we physically open it. What's interesting is that this interaction doesn't seems take much advantage of gesture-based interaction. Is it really easier to make a gesture than to physically pull out the drawer? The shape of the drawer provides a clue that it can be pull out. According to Don Norman, it gives a strong affordance to indicate how the drawer should be operated. However, with gesture-based interaction, the design principle proposed by Don Norman doesn't fit well. By given a drawer, it is hard for a user to guess which gesture exactly he/she should make to open it. They will either need to be trained or try several times to find the right gesture.

Other designers decide to create their own rules. For example, in the Wooden Mirror project created by Daniel Rozin, the wooden chips flip when sense the presence of human. Alloplastic Architecture project (Farahi Bouzanjani, et al, 2013) is another example where a tensegrity structure always follows the dancer's movement.

Whether following the existing rules or make new rules, the general approach is that we as designers pre-define the rules and users are asked to adapt into these "cookie-cutter" gestures. To make the adaptation process easier for user, architects use method like gesture elicitation study to increase the discoverability. (Wobbrock, Jacob O., et al, 2005) The goal of this widely used method is to extract appropriate gestures, which are easy to performance, memorable, and reliable. (Morris et al, 2014) Although architects try hard to make the adaptation process as easier as possible, it's still the user who is expected to learn. But if we look back to the very first quote from Nicholas, a responsive architecture is capable of recognizing the activities of users as well as responding to their needs.

As a result, I envision the future of human building interaction lies into the second approach where the relationship emerges from the interaction so that it is the user and building both learn from each other. In other words, there are two steps in the human building interaction. Human as user, expresses his needs, and architecture as provider, recognizes the intention and responds accordingly. In this case, both user and architecture need to learn. The user learns how to better express himself, while architecture learns how to better understand the user's needs.

Furthermore, I propose a 3-dimensional theoretical framework which includes predictability, discoverability and dimensionality, that can help designers to classify as well as design better gesture-based interactions

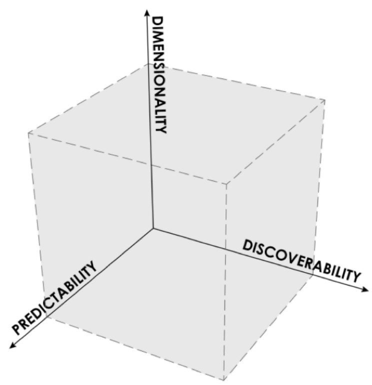


Figure 3: Three design constraints

Predictability refers to the ability of the system to behave in the way expected. It is certainly crucial as it is unpleasant to find that the gestures keep blending into each other. Discoverability refers to user's ability to find the target gesture. Increasing discoverability is a popular topic that many designers in HCI focus on because high discoverability can bring lots of benefits to a gesture-based application, such as ease of use and efficiency. (Morris, M. R. et al, 2014,) Dimensionality, on the other hand, is a new axis I propose which refers to the complexity of relationship between gesture and desired effect. It can be influenced by the number of gestures that are involved in the interaction as well as the complexity of these gestures. For example, an automatic door is considered as a low dimensional interaction because there is only one relationship which is moving the body towards the door. In the case of using multiple simple gesture like move hand from left to right, it is considered as medium dimensional interaction because the the increasing of gestures raises the complexity of the relationship. Using multiple complex gestures, such as walking, jumping, is considered as high dimensional interaction.

These three constrains are interdependent. If the predictability is maintained, high dimensional interaction usually means low discoverability and vice versa.

1.1 Related Work

Existing works shown that in this 3-dimensional relationship, most designers chose to keep the predictability at a relatively high level while trading dimensionality off with discoverability.

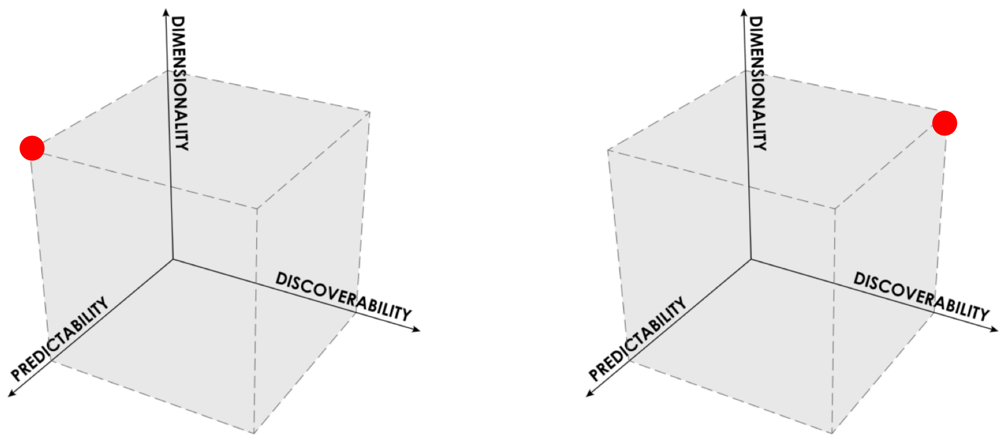


Figure 4: Design approach in HCI: High predictability and complexity(left)
Figure 5: Design approach in HCI: High predictability and discoverability(right)

In extreme cases, an application with the goal of security and safety concern can require high dimensionality with low discoverable interaction. (Figure 4) A gesture-based screen lock (Jeongyun, 2014) is an example which uses a specific gesture created by user to unlock the screen. In order to ensure only the creator of the gesture can unlock the screen, the gesture is better to be complex enough for it not to be discovered by other people. Project Soli is another example developed by Google ATAP. It is a new sensing technology that uses radar to detect subtle finger gestures and translate it to control a virtual interface. Because the

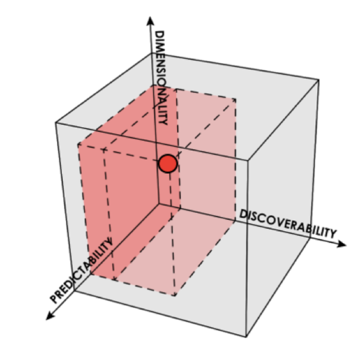


Figure 6: Trading discoverability with dimensionality

goal is to make this touchless interface mimicking the similar physical interface, the design group wants to make sure the predictability is maximized in this application. To accomplish more tasks, discoverability is traded off with dimensionality.

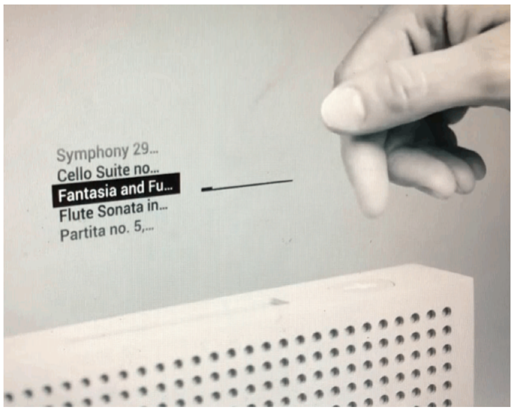


Figure 7: Project Soli, Interactive Sensor (Google ATAP, 2019)

Figure 8: Alloplastic Architecture (Farahi Bouzanjani, 2013)

In other cases, discoverability is typically more important. (Figure 5) For example, the alloplastic architecture (Farahi Bouzanjani, B. et al, 2013) is an adaptive tensegrity structure which intends to establish a scenario where a dancer can dance with the structure such that it reacts to her presence without any physical contact. Since the designer chose to maintain predictability as well, dimensionality was traded off with discoverability. The interaction itself is highly discoverable. But it is also very simple and the direct connection between the presence of human body and movement of the structure leaves user with nothing to discover after a few minutes. There are many installation projects designed by architects such as Spawn (Ramsgard-Thomsen, 2005), Spider (McKinney et al) that do the same thing as the Alloplastic Architecture does. The designers simplify

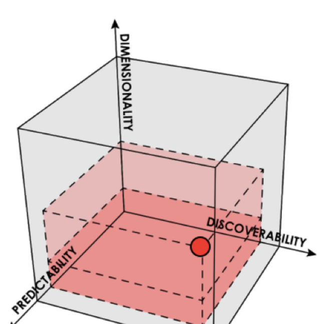


Figure 9: Trading dimensionality with discoverability

the interaction by minimizing the data that they collect, e.g. only collect occupancy or visual data, and limiting interaction type, e.g. the system only mimic user's gesture, in order to increase the discoverability. Although all of the projects claim that they create an interesting and attractive interaction between user and space, one can argue that when users said they were interested in the project, what they actually meant is that they were interested in the technology itself instead of the interaction. It is true that the interaction is predictable, and the gestures are discoverable. However, the connection between gesture and the movement generated by the system is too simple. Most users can figure it out within several seconds. Furthermore, the installation seems to be controlled by the user. In fact, the user has no control on the space movement at all. They can't control when to actuate and stop the movement. They can't choose what movement to have performed in response. Imagine if it is not a gesture-based interaction, but a conversation that happens between two people, with one person always repeating the other person's word, this conversation might end in a few seconds. In terms of the goal, it could be argued that these installations were designed with more concern regarding system recognition issues, than the end-usability of such gestures (Wobbrock, Morris, & Wilson, 2009). By over-simplifying gesture data and reduce the dimensionality of an interaction, we lose the information that conveys with the human gesture. A non-human object can trigger those interactions as well.

Figure 10: Mindful Photons (Zhao, N. et al, 2015)

Therefore, we might ask the question, is it always worth it to maximize the predictability? Why can't we maximize both discoverability and dimensionality and make it a richer communication between human and building?

Mindful Photons is a context-aware lighting system which aims to use google glass and infrastructure sensors to detect human activities and use them to control lighting. (Zhao, N. et al, 2015) The idea is that when the user is working, the eye focus various with the activity. When the user is relaxing, the eye focus does not change significantly.

In this project, the environment is "trained" to be aware of different using patterns. When the system recognizes certain activity, the lighting condition will be changed to fulfill the user's needs.

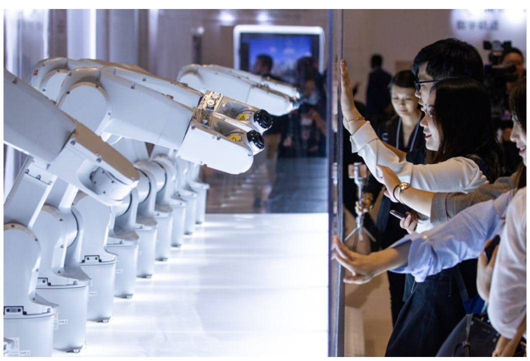


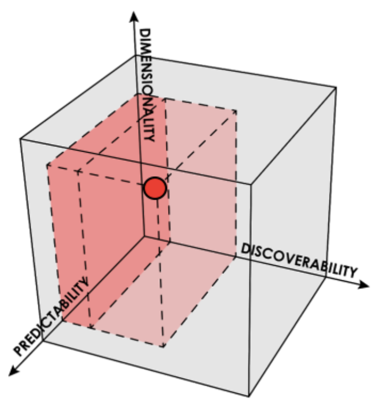
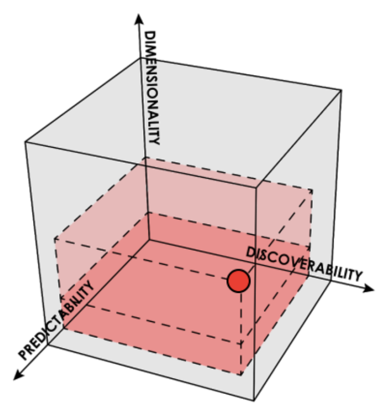
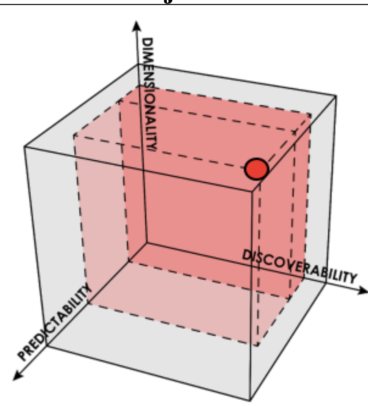
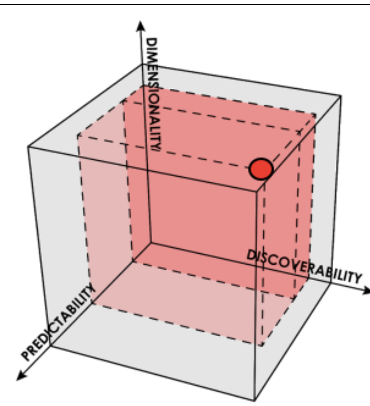
Figure 11: Manus (Madeline Gannon. et al, 2018)

Manus is another example which is an interactive robotic installation developed by Madeline Gannon who is a passionate researcher aiming to invent better ways to communicate with machines. (Madeline Gannon. et al, 2018) After being trained, her robot is able to guess the intention behind a human gesture and adjust itself to respond to it. For example, when the robots notice the presence of human, they look towards it before moving.

Overall, when focus on functionality, we want to make sure to maximize predictability. In the case of user learn, discoverability is traded off with dimensionality to fulfill more difficult tasks. Designers usually use gesture elicitation study to mitigate the risk of low discoverability. In the case of machine learn, both discoverability and dimensionality can be high and with the learning process, the predictability will be increased gradually.

When focus on Expression, the willingness to engage in a spatial interaction exists when people feel that they are “connecting” with the system (predictable) and are always able to discover the gesture command intuitively (discoverable). But more importantly, the interaction is better be diverse and rich enough for them to explore with. In the case of user learn, if we still keep the predictability and expect a relatively high discoverability like the example of Alloplastic, the dimensionality will be very hard to raise. However,

Table 1: An overview of related work under proposed framework

	Functionality	Expression
User Learn	 **Project Soli**	 **Alloplastic Architecture**
Both Learn	 **Mindful Photons**	 **Manus**

by having the architecture and user both learn from each other, we will be manage to keep the dimensionality and discoverability high and gradually increase the predictability by learning process. (Table 1)

1.2 Thesis Contribution

Japanese architect Sou Fujimoto once described architecture as a cave which refers to a naturally occurring and pre-existing condition. It exists independently. “If a human decides to occupy a cave, he or she must assimilate the lives to that which is already there.” (Brandon Donnelly, 2016) With all the physical interactions in architecture which limits our ability, people are always adapting themselves. If a door knob is designed in a round shape, people will adapt to turn the knob even when they think the door should be pushed to open.

However, a responsive architecture which will learn from the user and respond accordingly is more like the future that I envision gesture-based interaction could bring. Ideally, all the heavy training and adapting process are put on the back end of the system. The user will be able to use any gesture that makes more sense to them at the moment to interact with the architecture. Multiple gestures can be used to trigger the same interaction.

The thesis makes two contributions:

First, I propose a 3-dimensional theoretical framework that can help designers to classify and design better gesture-based interactions.

Second, I demonstrate a proof-of-concept by developing an interactive prototype and raising a discussion on both scenarios of user learn versus user and architecture both learn. By observing two groups of users interact with a customized art installation, I found that the willingness to engage in a spatial interaction exists not only when people feel that they are "connecting" with the system (the system is predictable) and are always able to discover the gesture command intuitively (the system is discoverable). But more importantly, the interaction is better be diverse and rich enough for them to explore with. The results of my user study confirmed the potential of such an interactive system to inspire its user and increase the willingness and curiosity among random users.

At the end of this paper I demonstrated an implementation for a choreography tool. By having this interactive art installation learn human gesture ahead of time, a dancer was able to "communicate" with it and created a live performance.

1.3 Research Focus

This research chooses to focus on human-center view which represents people's experience, feelings and perceptions on a gesture-based interaction. One important reason is that we could not make hasty conclusion that the most advanced technology was equal to the trend of future, but we can argue that a discoverable, entertaining, pleasant interaction can certainly find its way in the future implementation. For example, the speech dialing application, which based on the advance speech recognition technology (Dobler, S., 2000) has not been in widespread use. The reason is mostly usability concerns such as the user does not want people around know to whom he is dialing, or the phone won't recognize the voice dialing command because of the noisy environment. The trouble caused by such unstable and error-prone condition will certainly make the user feel frustrated. And one can argue that pressing keys is easy enough for people to make phone calls. As a result, to make this brilliant technology widespread, a study on how to make speech dialing more acceptable and intuitive is more critical than the technology itself. (Everhart, et al, 2005)

Therefore, this study focuses on the usability part of a gesture-based spatial interaction.

1.4 Research Questions

In order to investigate the previously mentioned issues regarding gesture-based spatial interaction, I answer the following research questions (RQ):

RQ1: How to make the user and architecture learn from each other in gesture-based interaction?

Objective: I demonstrate how the mutual learning process can be technically achieved by developing a machine-learning application and by designing and conducting a user study.

RQ2: What benefit could this mutual learning gesture-based interaction bring us? Where can we implement it?

Objective: I demonstrated an implementation for choreography and assess how the dancer feel during choreography.

CHAPTER 2: METHODOLOGY

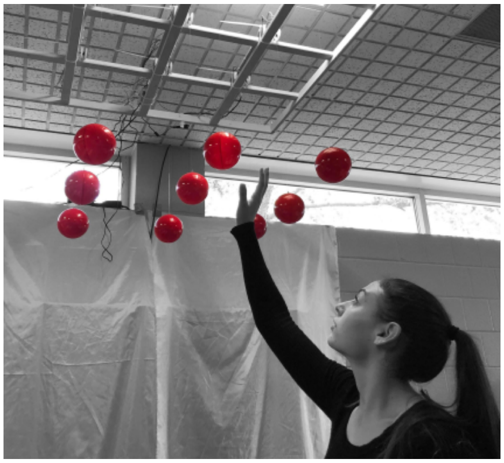


Figure 12: The NEST interactive art installation

In order to further discuss the difference between interactions that only ask user to learn versus architecture and user learn from each other, this study proposed an approach follows the procedure listed below.

I first design and create an interactive prototype named NEST. It is a 4’ by 4’ interactive art installation which is consist of nine stepper motors and programmed with Arduino Mega. Paired with a wrist attached sensor, this installation can be programmed to recognize human gesture and move accordingly.

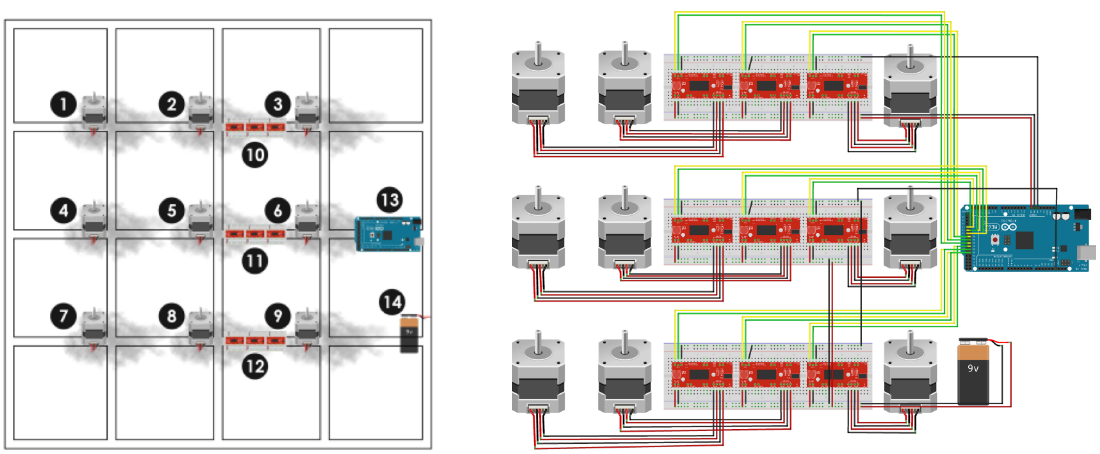


Figure 13: The NEST interactive art installation diagram

In the next step, I programmed the prototype to create two different modes. With the first mode, four pre-selected gestures derived from a traditional gesture elicitation study are manually paired with four move patterns. With the second mode, the installation is “trained” with 30 sample gestures for each movement and able to categorize more gestures based on a SVM algorithm. In order to collect gestures and program mode, I conducted a gesture elicitation study on 30 participants.

To make sure the result of user study won't be influence by gulf of execution, I choose to use a "data-glove based" method because most of "view-based" method is suitable in a controlled lab setting but does not generalize to arbitrary settings. If there are no high contrast stationary backgrounds and ambient lighting conditions, the recognition is very likely to make mistakes. Also, the machine cannot recognize the start and end points of meaningful gestures from continuous motion of the hands (Garg, P. et al., 2009). A "data-glove" with a switch button can ensure that every time the user makes the gesture, the computer can receive that data correctly.

I invited 30 people to propose gesture commands representing four different space movement patterns. This resulted in a list of 30 gestures for each one of the space movement patterns. A gesture classification system will then be developed. At the end of the process, a model is trained after data analysis with the ability to classify natural gesture and actuate space transformation. Once the model is trained, it will be used to control the mechanical ball system by recognizing natural human movement.

Furthermore, I conducted a user study in which I recruited 10 participants. 5 of them were asked to interact with the installation under 1st mode and 5 were asked to interact under the 2nd mode. This study compares the scenarios that user adapts versus user and architecture learn from each other. Later, I raised the discussion on how the development of an intelligent responsive system could eventually shape the future gesture-based interaction in architecture design.

CHAPTER 3: THE GESTURE ELICITATION STUDY

3.1 Apparatus and Referents

“NEST” is a 4’ by 4’ interactive art installation which aims to effectively show movement patterns with different complexity and examining the discoverability of gestures used to trigger these movement patterns.

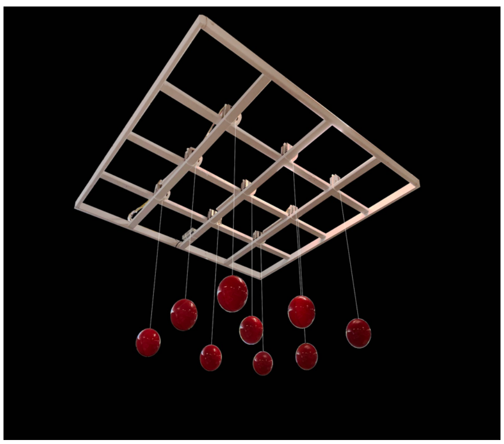


Figure 14: Nest Art Installation

Figure 15: Wrist attached sensor

Nine stepper motors are attached to a 3 by 3 wood grid from which hangs nine red balls. These balls can be programmed to move up and down smoothly. Four referents (space movement patterns created by NEST) were chosen from the pilot study to be presented to the participants including a Lean-slope shape, a Spiral shape, an X-shape and an L-shape. An Arduino controlled remote is designed to control this art installation.

I went through two prototype iterations prior to the official testing sessions. These iterations were informed by pilot testing with 5 participants. I identified the ideal space movement patterns by examining whether or not the type of shape can be clearly identified and described by participants. The goal is to have two patterns which are easy

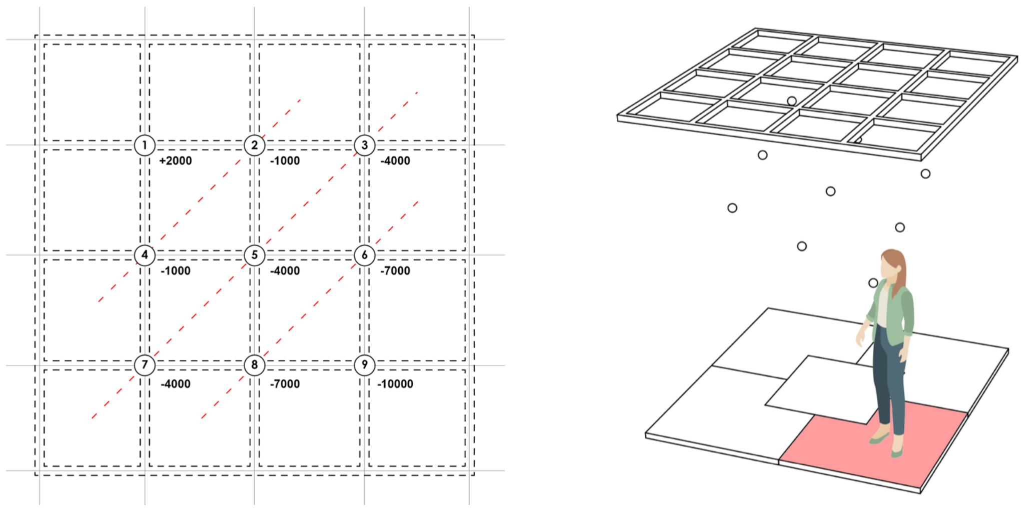


Figure 16: Lean-Slope shape

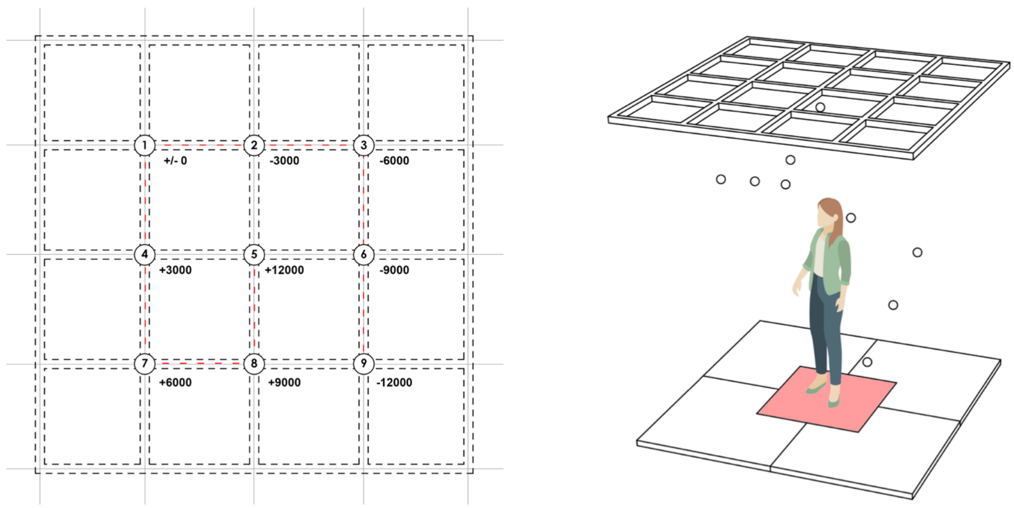


Figure 17: Spiral shape

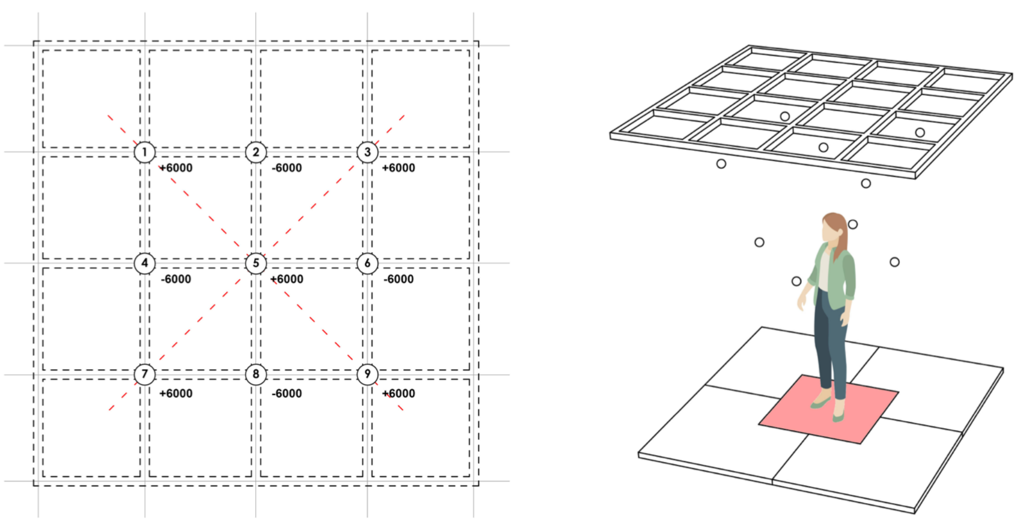


Figure 18: X-shape

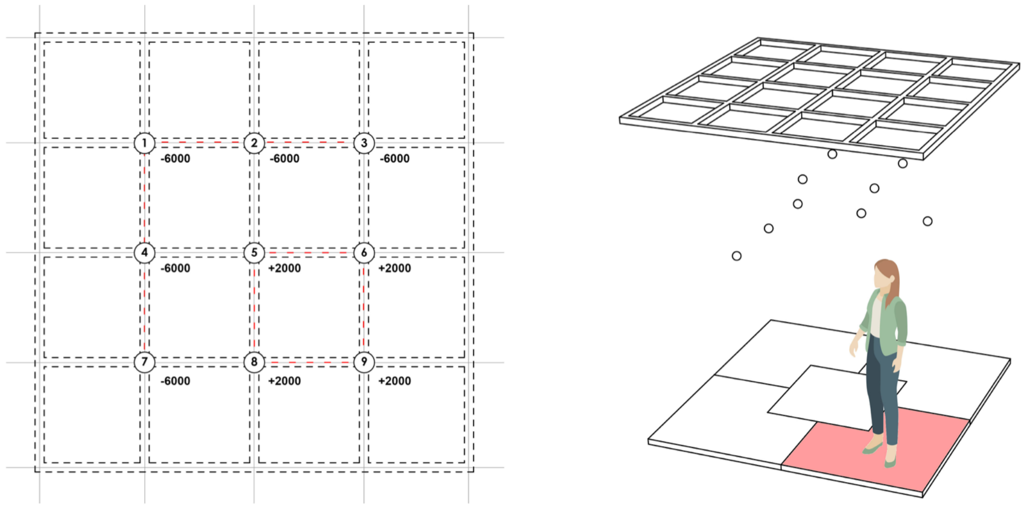


Figure 19: L-shape

to have consensus among participants versus two other patterns are difficult to be recognized.

In the pilot study, participants are presented with 8 space movement patterns created by NEST and asked to describe verbally with what they see. The two patterns (Lean-Slope shape, Spiral shape) with the highest agreement rate and two patterns (X-shape, L-shape) with the lowest agreement rate are chosen as the final patterns shown in the gesture elicitation study.

3.2 Participants and Tasks

30 participants were asked to propose gestures to trigger all four of spatial movement patterns. Participants must be able to physically make movements using full range of motion, have normal hearing and be able to understand the given tasks. All participants were right-handed and between 20-50 years old.

The participants were introduced with the purpose of this study first and sign the consent form. Before running the study, participants were given some time to familiarize with the the wrist attached equipment and its triggering method – the participant had to hold their thumb and forefinger together to start the recording process and release to end the recording process. Then, they were presented with all four of referents and were asked to propose a gesture to trigger each of the movement patterns. During the process, the participants were not allowed to ask questions regarding to the movement patterns. Participants were given as much time as they needed to propose gestures. Feedback from NEST were given to the participants by playing Wizard-of-oz. Once they were confident about their gesture proposals, the experimenter asked participants to reproduce the gesture 15 times for each in a random sequence so that it could be recorded by the Arduino controller in the format of 6-digit number and annotated by the machine learning algorithm (SVM). The experiment took around 15 minutes for each participant.

3.3 Results

The results were analyzed by traditional gesture elicitation study with the methodology of Wobbrock et al. and fed into a machine learning algorithm.

3.3.1 Traditional Gesture Elicitation Study

I measured consensus by calculating individual agreement rates for each referent with the methodology of Wobbrock et al. In my case, agreement rates vary between 0.07 (corresponding to the case with each participant proposing a distinct gesture for a given referent) and a maximum of 0.43 (perfect consensus between participants, all suggesting the same gesture for a given referent).

$$A = \frac{1}{|R|} \sum_{r \in R} \sum_{P_i \subseteq P_r} \left(\frac{|P_i|}{|P_r|} \right)^2$$

Figure 20: Agreement Rate (Wobbrock et al.)

The highest agreement rate was obtained for the “Lean-Slope shape” (0. 43), for which 13 participants out of 30 proposed right-hand movements to from the highest point to the lowest point.

The lowest agreement rates (0.2) were obtained for abstract space movement patterns, “X-shape”. 8 different gestures were proposed in total. 6 participants out of 30 agree on the gesture using right hand to draw a 2D “X” vertically.

Four gestures were picked from this study for triggering the four-space movement pattern.

Thinking time

Thinking time is defined as the time each participant spend on proposing a gesture. Although the thinking time spent here doesn’t relate directly to the discoverability of final product, it seems related to the agreement rate. And the agreement rate is a critical matrix that influence the discoverability. The higher the agreement rate is, the higher the discoverability will be.

By comparing the thinking time with agreement rate, it shows that the more time participants took to think about gestures, the less agreement resulted.

Participants took around 20 seconds to propose a proper gesture for “Lean-Slope shape” and “Spiral shape”. They spent a lot more in the “X-shape” (34 seconds) and “L-shape” (31 seconds) pattern. This is because the more time participants allocated to the task, the more different ways it can be translated into. “X-shape” and “L-shape” are just the label given to the two patterns by investigator for the ease of this paper. The participants in this study were not given any name or explanation of what the pattern is. Some of them think the “X-shape” is a big 2-D “X” which moves up. Some of them think the “X-shape” is a Jagger board. Because “X-shape” and “L-shape” were the two patterns with the lowest agreement rate in the pilot study, participants tended to spend more time on interpret what the pattern is. And it the more they think, the more different gestures were proposed.

Recall rate and Mistakes

Not like the traditional gesture elicitation study, this study was not only collecting data for human observer but also for a machine learning algorithm to learn from. As a result, each participant was asked to propose all four gestures first, recall and repeat them 15 times in a random sequence later. The recall rate is calculated as number of participants who successfully recall the gesture they proposed every time divided by the number of total participants. The recall rate for “Lean-Slope shape”, “Spiral shape”, “X-shape” and “L-shape” are 0.90, 0.86, 0.60, 0.66 respectively. Based on the statistic, best recalled gestures were found for simple movement patterns like “Lean-Slope shape” while lowest recall rates occurred for complex movement patterns “X-shape”. Another finding is that the gesture with higher agreement rate is more likely to be recalled correctly.

Mistakes can take place under three circumstances:

First, 12.5% of participants who made mistakes forgot the gesture right after he/she propose all four gestures. In this case, the participant usually asked more time or help from the investigator to recall the gesture. After being reminded, he/she can successfully recall all four of gestures throughout the study.

Second, 29.1% of participants who made mistakes forgot the gesture right after he/she propose all four gestures and after being reminded, he/she still can’t recall the gesture or recall the wrong gesture (a gesture that we referred to as a false positive) when asked repeating it 15 times.

Third, 58.3% of participants who made mistakes successfully recalled the gesture but mixed them when asked repeating.

Under second and third circumstance, the investigator needs to start the study from beginning and participants typically simplify their gestures.

3.3.2 Analyze Gesture Data with SVM

A Support Vector Machine (SVM) is a classifier algorithm which formally cluster data by hyperplane. In other words, given labeled training data (supervised learning), the algorithm outputs several hyperplanes which categorize new examples. The objective of using the support vector machine algorithm here is to search for the minimal enclosing spheres in a N-dimensional space (N = 6 - the number of features) that distinctly classify the gestures. These spheres each encloses a separate cluster of data points. In this case, all gesture data were labeled with space movement pattern number ("Lean-slope" – 1, "Spiral" – 2, "X-shape" -3, "L-shape" - 4) and participant number. For example, gestures proposed by participant "a" for space movement pattern "Lean-slope" will be labelled as 1a. Participant was asked to repeat each gesture 20 times as the algorithm requires plenty of samples in both groups. The data set was then divided into two group: ten of the examples were randomly chosen as the training set to train the model and the other ten examples were used for testing.

By selecting the proper kernel, q (the scale parameter) and C value (the soft margin constant), I was able to achieve 89.67% accuracy with the twenty gestures. All the gestures proposed in the elicitation study were then mapped back to space movement patterns.

There are several things to be notice here:

First, the accuracy on the test data set doesn't equal to the accuracy on the random data set. In other words, only when user propose a gesture that is similar to what's in the data set, the possibility of it to be successfully recognized is 89.67%. If a user proposes a new gesture which looks nothing like the gestures in the dataset, it will be tried to fit into one of the categories.

Second, I use OAA (One-against-all) in SVM (Vladimir Vapnik, 1995) which was first introduced as a method required unanimity among all SVMs: a data point would be classified under a certain class if and only if that class's SVM accepted it and all other classes' SVMs rejected it. While accurate for tightly clustered classes, this method leaves regions of the feature space undecided where more than one class accepts, or all classes reject. In the case of this data set, about 25% were unaccounted for. One method was proposed for improving the performance of OAA (Vapnik, 1998) involving the use of continuous values of SVM decision functions rather than simply their signs. The class of

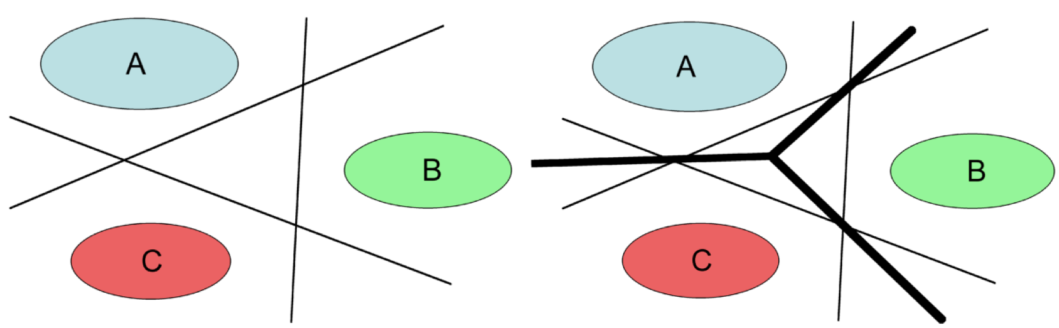


Figure 21: Diagram of binary OAA region boundaries on a basic problem (left)
Figure 22: Diagram of continuous OAA region boundaries on a basic problem (right)

a data point is whichever class has a decision function with highest value, regardless of sign. This appears to have a performance of approximately 96% correctly classified. About half of the data points unclassified by binary OAA were correctly classified by continuous OAA.

3.3.3 Advantages and Disadvantages

Low agreement rate

From what we see in the gesture elicitation study, agreement rates vary between a minimum of 0.2 and a maximum of 0.43, which is much lower than the average agreement rate in HCI. This result answers our second research question that it is harder to find a consensus on mapping a gesture with a space movement pattern. The more complex the space movement pattern get, the lower the agreement rate is. For “X-shape”, the gesture with most agreement rate only gets 0.2 which basically means that there is no agreement among users. When everyone proposed a different gesture and it is extremely hard for a human observer to categorize. By using SVM, I have an algorithm looking at all of the gestures. We are not eliminating any gestures. They are all used to categorize a new gesture proposed in the user study. The database can grow with the use of the space. The more data we collected, the more accuracy we will get.

Refinement of gesture

Different from traditional gesture elicitation study, I ask participant to repeat each gesture 15 times in a random sequence, which result in the voluntary refinement of gesture by the participant themselves. Especially when presented with high complex space movement pattern (“X-shape” and “L-shape”), some participants proposed very creative gestures which them can’t recall during the process. These gestures will confuse the human observer in traditional gesture elicitation as well as decrease the accuracy when use a mathematical algorithm. In order to recall the gesture successfully, most of them choose to adjust their gestures to a simpler version and have more connection with the space movement pattern which is helpful for finding discoverable gestures.

Continuous feedback

Unlike a traditional gesture recognition algorithm which only looks for the target gesture, the SVM algorithm will try to label every gesture into the category that fits them the best. So, it can provide feedback every time when user proposes a gesture which feels like the space is guess what you mean by making the gesture. The more gesture we collect, the better the algorithm can capture the intention behind a new gesture.

Overlapping

The more gestures I collected for each space movement pattern, the more possibilities for overlapping to occur between gesture datasets. This will lead to reduction of predictability as people may perform the similar gesture and find out that it triggers different space movement pattern. Through the traditional gesture elicitation study, I found that same gestures were proposed with two different space movement patterns under two condition:

First, participants propose gesture in reflect of different interpretations of the space movement pattern which matches with each other. For example, there was two participants proposed the same gesture of moving their hand from lower level of their body to higher level of their body for both the “Lean-slope” shape and “Spiral” shape. The former participant explained that she proposed the gesture for “Lean-slope” shape because she wanted to mimic the shape of the lean slope. The latter participant explained that she proposed the gesture for “Spiral” shape because she saw the movement as something that goes up and she just wanted to propose a gesture that represent that movement. Both of the gesture makes sense, so I decided to keep both of them and see what people feel in the usability test.

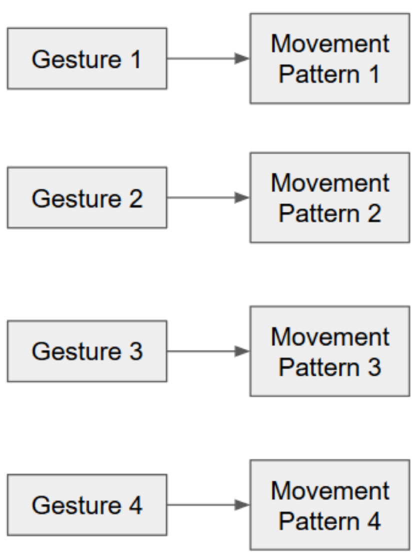
Second, participants propose informative gestures which is also called passive gestures. These gestures provide information about the speaker as a person such as their preference or culture and not about what the speaker is trying to communicate. For example, there was two participants proposed the same gesture of swiping their hand from left to right for both the “X-shape” and “L-shape” pattern. The former participant said he proposed gesture for “X-shape” because he couldn’t think of another gesture. The latter participant said she proposed gesture because she wanted to mimic the shape of an “L”. This gesture matched with 7 more gestures proposed for “L-shape” with other people and the reason given by the first participant is irrelevant. In this case, I choose to discard the gesture from the dataset of “X-shape”.

CHAPTER 4: USER EVALUATION STUDY

After collecting sample gestures, the NEST was then programmed with two modes:

In first mode, four gestures with the highest agreement rates in gesture elicitation study were assigned to the four space movement patterns. The participant can trigger the NEST only by making the target gesture. No feedback was provided when they made the wrong gesture.

In second mode, all the gestures proposed in the gesture elicitation study were categorized by SVM algorithm and assigned to its designated space movement patterns. The participant can trigger the NEST by exploring different gestures.



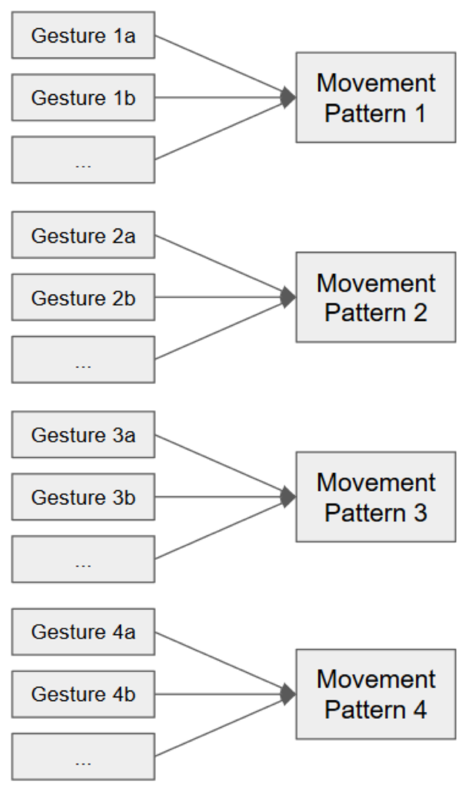


Figure 23: NEST programmed with one-to-one relationship (left)
Figure 24: NEST programmed with multiple-to-one relationship (right)

A usability study was conducted for assessing both methods when applying on a space interaction.

4.1 Participants and Tasks

10 participants were recruited to test the NEST. They were divided into 2 groups. Five of the participants were assigned to interact with the NEST programmed with the first mode and other five were assigned to the rest.

The participants were first shown the four movement patterns with NEST. The same glove-based sensor was then introduced to them. By using their gestures, participants were asked to trigger all four movement patterns within 5 minutes. If they successfully triggered all of the movement patterns with less than 5 minutes, the investigator stopped the timer and ended the study immediately. If they were not able to trigger all the movement patterns in 5 minutes, the investigator ended the study when time was up. The participants were then asked to fill in a pencil-based survey. All participants were welcomed to freely interact with the NEST after the study.

4.2 Survey Questions

1) From 1-10, how easy is it to interact with the installation? (1- very difficult, 10- very easy)

2) From 1-10, how well does the gesture recognition algorithm capture your intention behind gesture? (1- not very well, 10- very well)

3) If you were not able to trigger all four movement patterns, which one/ones was/were you not able to trigger?

4) From 1-10, how much did you adjust your gesture? (1- a little, 10 – completely changed)

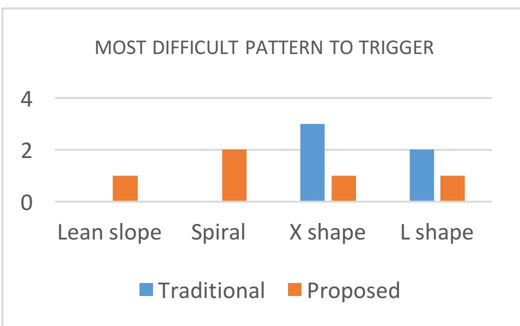
5) Among all four types of movement between balls, which one is the hardest one to trigger? Which one is the easiest one to trigger?

6) From 1-10, how interesting it is for you to interact with the installation? (1 – not interesting at all, 10 – very interesting)

7) Do you have any concerns during the interaction?

4.3 Observations and Discussion

The first observation is that two out of five participants who interacted with first mode successfully complete the task while all five participants with the second mode successfully complete the task in 5 minutes.



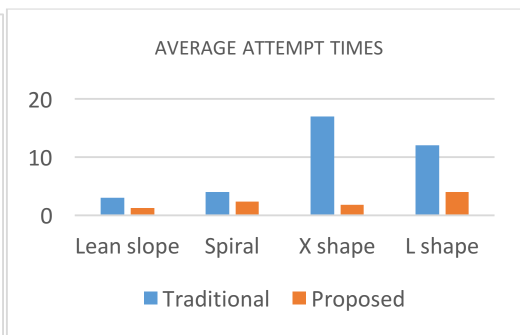


Figure 25: Most difficult pattern to trigger (left)
Figure 26: Average attempts spent on each movement pattern (right)

Among all participants interact with the NEST built with first mode, 3 participants thought “X shape” was the hardest pattern to trigger and 2 participants thought “L shape” was the hardest one. This result matches with agreement rate in gesture elicitation study. Obviously, the lower the agreement rate is, the less discoverable the interaction is. We can also find clues by looking at the average time that participants spent on triggering each movement pattern. Participants in the first group spent much more time on triggering complex space movement patterns (“X-shape”, “L-shape”) than simple space movement patterns (“Lean-slope”, “Spiral”) because of the low agreement rate. With no instruction and feedback, some participants even asked if the system was still working.

Table 2: Group 1 – Time spent on each movement patterns

	Lean Slope	Spiral	X-shape	L-shape	Total Time
Participant 1	18”	24”	3’24”	12”	4’18”
Participant 2	5”	17”	2’28”	1’40”	4’30”
Participant 3	25”	34”	4’1”	-	5’
Participant 4	17”	12”	3’2”	1’29”	5’
Participant 5	21”	26”	2’36”	1’37”	5’
Average Time	17.2”	22.6”	3’6”	-	

Table 3: Group 2 – Time spent on each movement patterns

	Lean Slope	Spiral	X-shape	L-shape	Total Time
Participant 1	5”	4”	5”	12”	26”
Participant 2	12”	2’24”	5”	33”	3’14”
Participant 3	7”	34”	7”	1’1”	1’49”
Participant 4	38”	27”	5”	13”	1’23”
Participant 5	11”	5”	12”	11”	39’
Average Time	14.6”	42.8”	6.8”	26”	

Here we can make the first conclusion that the traditional gesture elicitation study is good for low dimensional interactions which are easier to find consensus on the proposed gestures. A simple gesture controlled app will benefit from it. The pre-defined one-to-one relationship between target gesture and movement pattern makes the interaction very predictable. However, it is also this narrow gesture vocabulary that limits the discoverability. In HCI, when designing a gesture-based application with complex interaction types, this risk can be mitigated by having an instruction book and training the user with the target gesture. However, people experience architecture rather than use it. Architects can’t give everyone who is going to use the building an instruction book and we can’t expect that all users will use the space as we designed it. As a result, user will benefit more from a relationship that emerge from the interaction.

The second observation is that none of participant who were assigned to interact with the NEST programmed with second mode, triggered NEST movement one by one. They all start with the “Lean slope” shape and 4 out of 5 participants successfully triggered the movement with only one attempt. When they moved to the “Spiral” shape, 4 out of 5 of participants accidentally triggered the “X shape”. This is caused by two reasons: either this gesture has been proposed by participants in the gesture elicitation study for both space movement pattern “Spiral” and “X shape” or this gesture has never been proposed by any participant in the gesture elicitation study. In the first case, typically the participant would think for several seconds and said, “That makes sense”. Then they will try to find another gesture that can describe the “Spiral” better and be more differentiated from “X shape”. In the second case, because the gesture was never proposed before, the algorithm will try to fit it in one of the categories and the accuracy is not ensured. This is the part that the participants were confused because the connection between gestures and space movement pattern didn’t make any sense. Instead of proposing a new gesture, most of them tried to adjust their gestures slightly to see if they can trigger the “Spiral”.

In terms of discoverability, feedback in the first case is helpful because it not only helps the user understand which gesture triggers another movement pattern but also guide user to find a good gesture for the current movement pattern that they want to trigger. Feedback in the second case did no good to usability but confuse the user. However, I think this is something we can overcome by increasing the size of database and optimizing the machine learning algorithm.

The third observation is that participants in group 1 refused or minimized their free time to play with the NEST while participants in group 2 took 2-3 more minutes to interact with the system. In the survey, most participants in group 2 mentioned that they were curious about what other gestures can be used to trigger the movement and willingness to engage in the interaction is high in group 2 than group 1.

CHAPTER 5: DISCUSSION

By having the user adapting into the pre-defined “cookie-cutter” gestures, we can make a fully functional interaction. However, functionality is not the only thing user looking for in a gesture-based interaction with an architecture. Because there is no clear relationship between a gesture and a movement pattern, an interaction which will allow user to use any preferred gesture to operate the system can be powerful. We should not give an instruction book to every user. Instead, we provide the user with a tool where they can create their own interaction to communicate with the architecture. One scenario could happen in a museum setting where a functional gesture-based interaction might be interesting but if the visitor is unable to discover the right gesture to trigger it, without proper instruction, the interaction can be simply missed. However, if we can program the building to gradually learn the intention behind the gesture, user might accidentally trigger some irrelevant movement but feel curious about it. They then propose more gestures to explore the possibilities of the system.

Creating an intelligence environment which can adjust itself and guess the intention behind a gesture is definitely one route for gesture-based interaction to go in architecture design. Using Amazon Echo as an example, there’s voice command that you know will always work such as “Alexa, turn on the light!” But there’s also other sentences that you can try to trigger the same function. More important, what makes Alexa different from other home devices is that there are sentences with no fixed answer but will keep learning and updating to fit the user’s expectation such as “Alexa, how are you?” An application that can do the former is considered as a successful product but the application that can do the latter is the future.

REFERENCES

[1] Nahrstedt, K., Bajcsy, R., Wymore, L., Sheppard, R., & Mezur, K. (2007). Computational model of human creativity in dance choreography. Urbana, 51, 61801.

[2] Carlson, K., Schiphorst, T., & Pasquier, P. (2011). Scuddle: Generating Movement Catalysts for Computer-Aided Choreography. In ICCC (pp. 123-128).

[3] Enrica Fung, Kristof Corolla, 2018, Choreographed Architecture, Proceedings of the 23rd International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) in Hong Kong.

[4] Hansen, L. A., & Morrison, A. (2014). Materialising Movement–Designing for movement-based digital interaction.

[5] Meyer, C. M. (2006). Dance+ architecture: designing to inspire movement.

[6] Zhao, N., Aldrich, M., Reinhart, C. F., & Paradiso, J. A. (2015, November). A multidimensional continuous contextual lighting control system using Google Glass. In Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments (pp. 235-244). ACM.

[7] Benesh, R., & Benesh, J. (1977). Reading dance: The birth of choreology. International Specialized Book Services.

[8] Kaye, N. (2013). Site-specific art: performance, place and documentation. Routledge.

[9] Perusinovic, D. (2011). Choreography and Architecture (Doctoral dissertation, Carleton University Ottawa).

[10] Mantho, R., Crotch, J., & Horner, M. (2010). Social spatial genesis: activity centered space making.

[11] T. Fukuda, W. Huang, P. Janssen, K. Crolla, S. Alhadidi (eds.), Learning, Adapting and Prototyping, Proceedings of the 23rd International Conference of the Association for Computer-Aided Architectural Design Research in Asia (CAADRIA) 2018, Volume 1, 101-110.

[12] Marey, Etienne-Jules. Du mouvement dans les fonctions de la vie. [Movement in the functions of life].Paris:G. Baillière, 1868.

[13] Vulinovich, R. (2009). Performance space and public movement (Master's thesis).

[14] Zhang, X., Dekel, T., Xue, T., Owens, A., He, Q., Wu, J., ... & Freeman, W. T. (2018, October). MoSculp: Interactive Visualization of Shape and Time. In The 31st

Annual ACM Symposium on User Interface Software and Technology (pp. 275-285). ACM.

[15] Behnaz, F. B. (2016). Alloplastic architecture. Interactive architecture. Princeton Architectural Press, New York.

[16] Zoran, A., Shilkrot, R., & Paradiso, J. (2013, October). Human-computer interaction for hybrid carving. In Proceedings of the 26th annual ACM symposium on User interface software and technology (pp. 433-440). ACM.

[17] Bonnemaison, S., & Eisenbach, R. (2009). Installations by Architects: Experiments in building and design. Princeton Architectural Press.

[18] Thomsen, M. R., & Brown, C. (2004, August). Conceiving embodiment: the dance-architecture of spawn. In ACM SIGGRAPH 2004 Sketches (p. 114). ACM.

[19] Bard, J. D., Blackwood, D., Sekhar, N., & Smith, B. Decorative Robotic Plastering.

[20] McCullough, M. (1998). Abstracting craft: The practiced digital hand. MIT press.

[21] Gannon, M. REVERBERATING ACROSS THE DIVIDE.

[22] Alaoui, S. F., Carlson, K., & Schiphorst, T. (2014, June). Choreography as mediated through compositional tools for movement: Constructing a historical perspective. In Proceedings of the 2014 International Workshop on Movement and Computing (p. 1). ACM.

[23] Camurri, A., Hashimoto, S., Ricchetti, M., Ricci, A., Suzuki, K., Trocca, R., & Volpe, G. (2000). Eyesweb: Toward gesture and affect recognition in interactive dance and music systems. Computer Music Journal, 24(1), 57-69.

[24] Bermudez, B., Delahunta, S., Marijke, H., Chris, Z., Bevilacqua, F., Alaoui, S. F., & Gutierrez, B. M. (2010). The double skin/double mind interactive installation. The Journal for Artistic Research.

[25] Calvert, T., Wilke, W., Ryman, R., & Fox, I. (2005). Applications of computers to dance. IEEE computer Graphics and Applications, 25(2), 6-12.

[26] Jacob, M., & Magerko, B. (2015, June). Viewpoints ai. In Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition (pp. 361-362). ACM.

[27] Wallis, M., Popat, S., McKinney, J., Bryden, J., & Hogg, D. C. (2010). Embodied conversations: performance and the design of a robotic dancing partner. Design Studies, 31(2), 99-117.

[28] Singh, V., Latulipe, C., Carroll, E., & Lottridge, D. (2011, November). The choreographer's notebook: a video annotation system for dancers and choreographers. In Proceedings of the 8th ACM conference on Creativity and cognition (pp. 197-206). ACM.

[29] Lynn, G., & Kelly, T. (1999). Animate form (Vol. 1). New York: Princeton Architectural Press.

[30] Thomsen, Mette Ramsgard, and Carol Brown. "Conceiving embodiment: the dance-architecture of spawn." ACM SIGGRAPH 2004 Sketches. ACM, 2004.

[31] Bradford, J. H., & Côté-Laurence, P. (1995). An application of artificial intelligence to the choreography of dance. Computers and the Humanities, 29(4), 233-240.

[32] Brand, M., & Hertzmann, A. (2000, July). Style machines. In Proceedings of the 27th annual conference on Computer graphics and interactive techniques (pp. 183-192). ACM Press/Addison-Wesley Publishing Co..

[33] Nielsen, M.; Störring, M.; Moeslund, T.B.; Granum, E. A Procedure for Developing Intuitive and Ergonomic Gesture Interfaces for HCI. In Gesture-Based Communication in Human-Computer Interaction, Proceedings of the 5th International Gesture Workshop, GW, Genova, Italy, 15–17 April 2003; Springer: Berlin/Heidelberg, Germany, 2003

[34] Vogiatzidakis, P., & Koutsabasis, P. (2018). Gesture Elicitation Studies for Mid-Air Interaction: A Review. Multimodal Technologies and Interaction, 2(4), 65.

[35] Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE, 77(2), 257-286.

[36] Rahman, A. S. M., Hossain, M. A., Parra, J., & El Saddik, A. (2009, October). Motion-path based gesture interaction with smart home services. In Proceedings of the 17th ACM international conference on Multimedia (pp. 761-764). ACM.

[37] Pu, Q., Gupta, S., Gollakota, S., & Patel, S. (2013, September). Whole-home gesture recognition using wireless signals. In Proceedings of the 19th annual international conference on Mobile computing & networking (pp. 27-38). ACM.

[38] Ruiz, J.; Vogel, D. Soft-Constraints to Reduce Legacy and Performance Bias to Elicit Whole-Body Gestures with Low Arm Fatigue. 2015. Available online: https://hci.cs.colostate.edu/files/6614/3387/9908/paper2504.pdf (accessed on 8 September 2018)

[39] Morris, M.R.; Danielescu, A.; Drucker, S.; Fisher, D.; Lee, B.; Schraefel, C.; Wobbrock, J.O. Reducing legacy bias in gesture elicitation studies. Interactions 2014, 21, 40–45.

[40] Everhart, C. A., Gourd, D. V., Joshi, R. M., & Kefalos, J. (2005). U.S. Patent No. 6,845,251. Washington, DC: U.S. Patent and Trademark Office.

[41] Vapnik, V. (1995) The Nature of Statistical Learning Theory. Springer-Verlag, London.

[42] Vapnik, V. (1998). Statistical Learning Theory. Wiley-Interscience, NY.

[43] Jeongyun, H. E. O., Kim, D., Park, M., Lee, S., Jihye, H. A. M., Park, K., & Kim, S. (2014). U.S. Patent No. 8,756,511. Washington, DC: U.S. Patent and Trademark Office.

[44] Wobbrock, J. O., Aung, H. H., Rothrock, B., & Myers, B. A. (2005, April). Maximizing the guessability of symbolic input. In CHI'05 extended abstracts on Human Factors in Computing Systems (pp. 1869-1872). ACM.

[45] Norman, D. (2013). The design of everyday things: Revised and expanded edition. Basic books.

[46] Solanki, U. V., & Desai, N. H. (2011, December). Hand gesture based remote control for home appliances: Handmote. In 2011 World Congress on Information and Communication Technologies (pp. 419-423). IEEE.

[47] Brandon Donnelly, 2016, https://brandondonnelly.com/2016/04/22/nest-or-cave/

[48] Thiruvengada, H., Laberge, J., Foslien, W., Derby, P., Putrevu, S., & Vargas, J. (2013). U.S. Patent Application No. 13/367,015.