Touchless Gestural Interaction with Wizard-of-Oz: Analysing User Behaviour

Martin Henschke
Australian National University
Canberra, Australia
martin.henschke@anu.edu.au

Tom Gedeon
Australian National University
Canberra, Australia
tom.gedeon@anu.edu.au

Richard Jones
Australian National University
Canberra, Australia
richard.jones@anu.edu.au

ABSTRACT
In many gestural interfaces, the gesture set is developed or trained by real users, but many use gestures with unchanging definition that do not account for variation between different users or performances. Over time, if the user performs gestures different, the definition should evolve to accommodate these changes. We performed a Wizard-of-Oz experiment with a user-defined gesture system to determine if participant's gestures changed over repeated performances, and when and why these changes occurred. The results showed that although the definitions provided changed in a unique way for each participant, most reported their gestures as becoming simpler and less difficult to perform over time.

Author Keywords
"Touchless Interfaces", "User Experience", "Wizard of Oz"

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION
Touchless gestural interfaces provide a potential way for users to interact with computers and machines in a manner they find more intuitive, easy or pleasing (Wachs et al., 2011). Although touchless gestural interfaces are becoming increasingly sophisticated with improved hardware and recognition systems, these systems remain challenging and fatiguing to use (Cabral et al., 2005, Hincapié-Ramos et al., 2014). One possible reason is many gestural interfaces rely on users performing gestures consistently, to ensure the recognition is accurate. Many systems have users train before or during use to account for this (Kratz et al., 2007, Wright et al., 2011). However, such implementations do not account for how the user will change performance after the training period is over. Systems that rely on a predefined set of gestures also face the issue of the gestures being unintuitive or not uniformly understood by all users (Malizia and Bellucci, 2012). This paper therefore looks to explore if users are inclined to change the way they self-defined gestures after repeated usage, and make observations in what ways these changes manifest. Our goal is to determine if any patterns or common behaviours are apparent that may aid in designing gesture-driven systems for general use. While there has been a reasonable amount of research into how users prefer to select and perform gestures for a system, our interest is focussed on how users behave once a gesture has been selected. Specifically, we wanted to determine if users are inconsistent in how they perform gestures or how they prefer to perform it after some exposure to a system.

OzCHI '15 , December 07 - 10 2015, Melbourne, VIC, Australia
Copyright © 2015 ACM 978-1-4503-3673-4/15/12... $15.00
http://dx.doi.org/10.1145/2838739.2838792.

Related Work
In designing gestures for an interface, selected gestures typically should be appropriate to the task, easy to learn, perform and remember (Nielsen et al., 2004). The taxonomy of the interface can be defined solely by the designer, though systems like those reported in Wobbrock et al. (2009) have users design gestures without guidance. In developing gestures, users typically prefer iconic, or 'miming' gestures (see Cassell (1998)) that represent the usage of a tool or mimic a real-life activity (Grandhi et al., 2011). Examples of systems that make use of trained gesture sets can be seen in Kratz et al. (2007) with Wiimote controllers and in (Wachs et al., 2011, Cabral et al., 2005) using hand tracking.

This experiment makes use of the 'Wizard of Oz' design, in which a participant interacts with an apparently intelligent system, unaware that its responses are provided by a human operator referred to as the *wizard*. Wizard of Oz experiments are often run in natural-language interfaces (Dahlbäck et al., 1993, Fiedler and Gabsdil, 2002). In this experiment however, the wizard is performing adaptive gesture recognition, as opposed to an automated system (Beringer, 2002). The main reason for this choice was having a more interpretive and flexible human operator allows participants to perform gestures in the manner that they desire rather than having to change their gestures to fit within the limitations of the recognizer. It is hoped removing this inhibition will allow us to better observe the changes in performance over repeated trials. The use of a Wizard of Oz trial over a transparent human-human experiment of a similar nature may also fail to capture the idiosyncrasies of human behaviour in front of a recognizer, due to the disconnect in understanding between system and user (Fothergill et al., 2012).Other reasons for choosing this approach include greater control by the experimenter over how the experiment is run, and no requirement for a training

period, which may change how the user performs in the longer run.

EXPERIMENTAL DESIGN

System

The system we used to perform this experiment had two distinct parts: the view, which is operated by the experimental participant and the controller, which is operated by the experimenter.

The user view displayed a 3D scene that participants were able to interact with. The scene consisted of a series of doors with locks in random positions on each door. Participants were given the goal of summoning a key, and manipulating it using gestures to unlock and open each door. Once each of the six doors they encountered were unlocked and opened, the trial was complete.

The participants were able to interact with the scene using eight separate gestures; 'summoning' the key to make it appear in the scene, rotating the key to change the direction it is facing, turning the key to open the lock, opening an unlocked door and four distinct gestures for moving the key up, down, left and right a discrete distance. The goal for each step of the experiment was to summon the key, rotate and move it to the correct position next to the lock, turn it to unlock the door and finally open the door. At the beginning of the experiment, the participant was asked to assign each of these interactions a gesture, in the form of some body movement that could be recorded in a 2-5 second window.

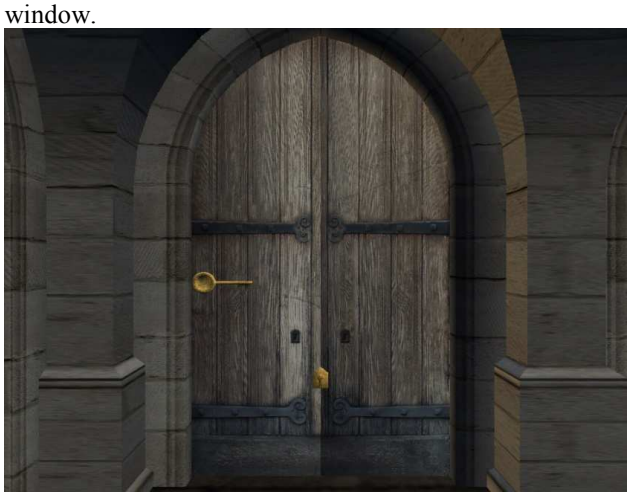

Figure 1: The participant view in the experiment

The experimenter, or 'wizard', viewed the system from a separate display, hidden from the participant. This view showed a live camera feed and repeating video loops of the each gesture the user has defined, which were unlabelled so the wizard was not explicitly aware or what each gesture meant. At the beginning of an experiment, the wizard watched the feed and created the recording for each gesture the participant defines. Once all gestures had been defined, the wizard watched the live camera feed. When the participant performed an identifiable gesture, the wizard selects the appropriate video loop, prompting the system to perform the designated action.

Experiment

At the beginning of the experiment, each participant was briefed on how the system worked, including each of the interactions they could perform and the goal of the experiment. The participant was then introduced to the system, and began the process of defining gestures and opening the set of six doors, which constituted a single trial. Each participant completed four separate trials. At the end of the experiment, participants filled out a short questionnaire reporting on any fatigue they experienced during the experiment and how they perceived the change in their own gesturing, as well as how regularly they felt the system reported false positives (recognizing a gesture when none was conducted) and false negatives (recognizing the wrong or no gesture).

Due to the difficulty in measuring the 'simplicity' or 'complexity' of a gesture, a number of measures were taken. A Microsoft Kinect was used as the camera, which in addition to capturing a camera feed, also constructed a skeletal model of each gesture definition recorded. In processing the data, the start and end frames of each gesture were specified by the experimenter using a custom program that superimposed each gesture in a trial (Figure 2). Where gesture definitions were the same, the distance of movement between the user's final position when performing the gesture and their default rest state was then recorded. In the event a definition was changed sufficiently to make it incomparable, the nature and characteristics the change was recorded.

Hypotheses

It was expected that most users would need to redefine gestures near the start of usage if they found the provided definition to be ambiguous in the context of other gestures or difficult to perform frequently. Once a gesture was memorized and comfortable to perform, we did not expect users would need to redefine it.

The second expected behaviour was that for a gesture that was not redefined, later performances and definitions of the gesture would be more compact and less strenuous to perform than earlier definitions. We define this in the results as a gesture that either uses fewer parts of the body to perform (e.g. using just the lower arm when previously defined using the whole arm), or less movement between the gesture's starting position and its terminating position.

RESULTS

A total of fifteen volunteers participated in the experiment, 13 males and 2 females with a mean age of 21. All but one of the participants are undergraduate students studying Computer Science or IT as part of their degree. Nine reported familiarity with gesture control devices like Nintendo Wiimotes, but only five were familiar with touchless gestural systems like the Kinect.

Questionnaire Results

The mean reported fatigue at the end of the experiment, on a scale from 1 to 10, was 2. Only one participant reported considerable fatigue with a score of 7, and attributed this discomfort to one large gesture that caused shoulder pain. Four participants reported the recognition of false positives during the experiment, two of which mentioned they only occurred once or twice. False

negatives were reported by 10 of the participants. Of those, 4 participants suggested the reason was a lack of distinction between two separate gestures that lead to an ambiguity, 1 reported it being caused by the gesture being too small or difficult to recognize, 1 reported it being because the gestures became too loose and lazily performed over time, and the remaining 3 did not suggest a reason.

Of the participants, 14 of 15 reported their gestures changing over the course of the experiment, for a diverse set of reasons. To compare these with the hypothesis, the language used in each response was categorized as meaning 'simpler' gestures or 'more complex' gestures. Of those, 3 reported their gestures becoming more complex or difficult to perform, 10 reported their gestures becoming simpler and easier to perform and 1 reported having done both. A variety of terms were used to describe how gestures were simplified; these include 'compact', 'smaller' and 'closer to the body' (3 reports) which is interpreted as making less movement overall and performing closer to a rest state, 'faster' (2 reports) or taking less time to perform the gesture, 'lazier' and 'efficient' (3 reports), or using simplified and less demanding movements to perform the gestures. The expression 'flowing' and 'streamlined' (3 reports) as used is somewhat ambiguous but suggests either less distinct steps within a gesture, or gestures that facilitate being performed in sequence easier. The participant that reported a high fatigue score explicitly described the simplification of the gesture as a means to combat the discomfort.

In making gestures more complex, descriptions included using both hands rather than one (1 report) and 'exaggeration' (1 report), making the gestures movement larger and easier to recognize. Two participants also reported changing the gesture for the purpose of 'novelty'; several gesture definition they provided would be highly impractical when using the system more seriously, such as jumping and ducking to move the key up and down, or performing a kick to open the door.

Two participants also used the similar terms 'differentiated' and 'distinct', which doesn't indicate simpler or complex directly. Both participants also reported false negatives, so it is expected this refers to trying to minimize these errors.

Observations

In analysing the gestures performed qualitatively, we define a 'change' in definition being a movement that has one or more components that are distinct from the previous definition, either in the parts of the body that are used to perform it or the direction in which one or more components of the movement takes place. Changes were categorized as either 'simpler' or 'more complex' according to the criteria participants described in the questionnaires (see above). In some instances, a new gesture was clearly complex or had more extreme motion in how it is performed, while in other instances a gesture was different without a clear way to categorize it according to our criteria. These instances are referred to as 'neutral' changes. A common example of this was changing the arm used to perform a gesture, or the direction it was performed. Finally, instances existed where a gesture definition was the same as one provided in an earlier trial; these are identified as reversions. Each participant had 3 chances to redefine a gesture (at the end of the 1st, 2nd and 3rd trials), for a total possible 24 redefinitions per experiment.

The mean number of redefinitions that occurred over all participants in an entire experiment was 3.6 changes. Of those changes, 58.2% were changes to a more complex definition, 14.5% were to a simpler definition, 23.6% were reversions to a previously defined gesture and 3.6% were redefined neutrally. There were two notable outliers, one user redefined their gestures 13 times and the other 14 times over the course of the experiment.

Redefining to a more complex gesture made up the majority of all redefinitions. To determine if this was related to misinterpretations, or 'distinction' as reported by participants in the questionnaires, we compared the frequency of complex definitions with whether or not participants reported false negatives being an issue. Of the complex redefinitions we recorded in our experiment, 87.5% were defined by participants that reported false negatives. On average, a participant that reported a false negative therefore would redefine to more a complex gesture 2.8 times during the trial, while someone that did not report the problem would redefine 0.8 times per trial.

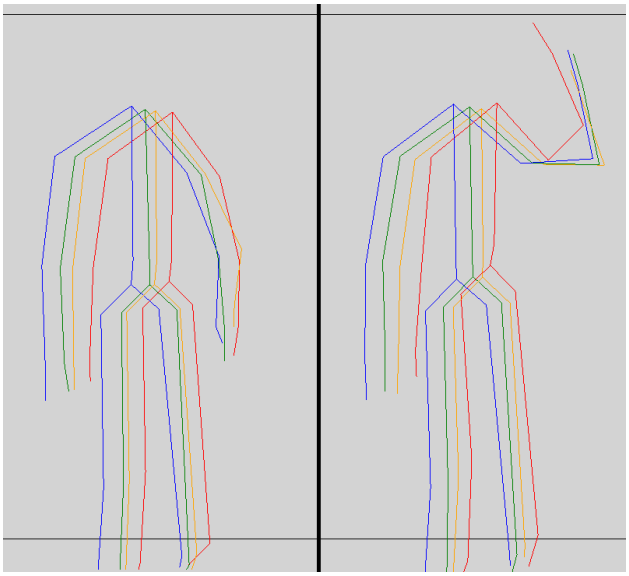

Figure 2: The starting and ending frames of the 'Move Up' gesture identified for one participant

There were several unusual occurrences during the experiment. In addition to the use of very impractical and difficult gestures for novelty or fun, one participant defined his gestures inversely, using a downward motion to move the key up and so on.

To find the difference in movement distance for similar gestures, the starting and terminating point were defined for each gesture, the start being defined as the point at which the body moves from a rest position, or moves to perform the defining action from the initial position the gesture was captured in (as several gestures were started

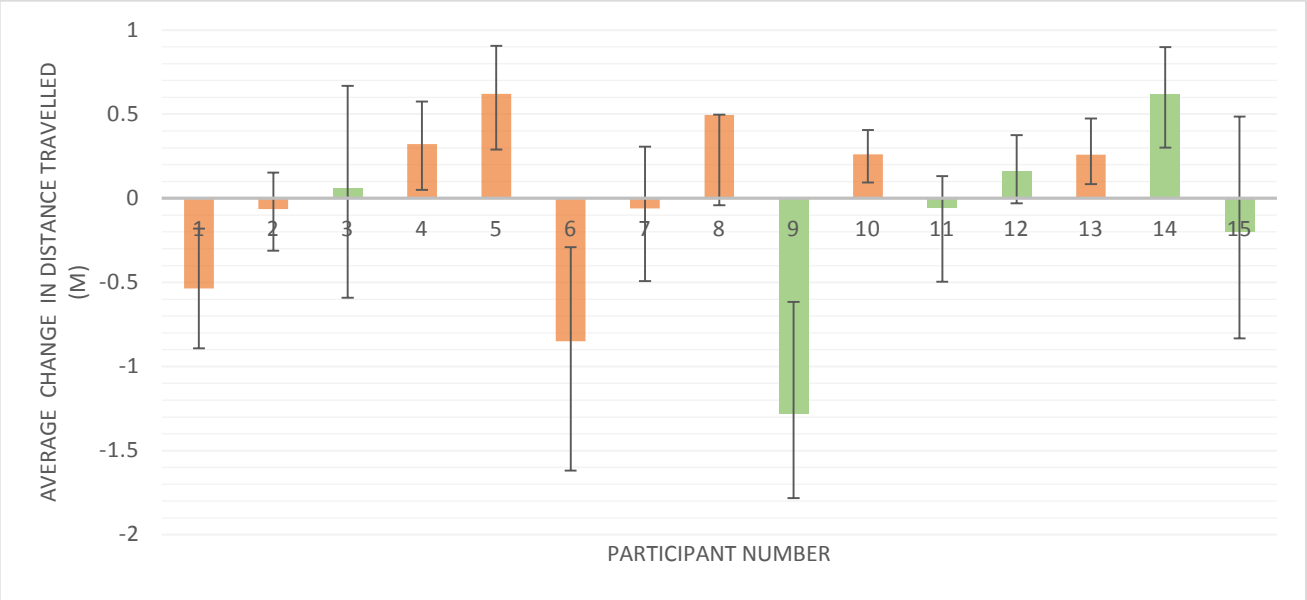

Figure 3. Variation in average distance travelled in performing gesture set over each trial

from the end of a previous gesture rather than rest). The terminal point of the gesture was typically picked as the most extreme point of the gesture in relation to the rest position. The duration of a gesture was either the time difference between the start point and the return to a rest state, or until the gesture had clearly terminated in the event the body did not return to a rest state. Gestures with similar definitions but dissimilar start and end points in terms of body position were considered too dissimilar to make reasonable comparisons and excluded. In addition, several gestures were very small or performed in ways that made the skeletal model produced by the Kinect with too high a margin of error to be considered.

With these results, the amount of user movement in a gesture was found by taking the last definition provided and finding the distance each relevant joint moved from the rest state to the final state, which provides a distance in centimetres. This result was scaled according to the minimum and maximum amount of distance the user made in any of their gesture definitions. The results are shown in Figure 3; negative values indicate the gesture became smaller over time, while positive represent more distance travelled in later gestures. Orange bars represent participants that reported false negatives in their experiments. The average distance is shown by each bar, with the error bars representing the inter-quartile range (IRQ) of difference in distance for all gestures the participant performed.

The figure shows a different finding than expected in the hypothesis. Although some users (with scores less than zero) did produce smaller gesture definitions at the end of their trial, most redefinitions were relatively insignificant and about an equal number defined later gestures with more movement rather than less.

DISCUSSION AND FUTURE WORK

Our experiment clearly demonstrated a strong tendency for users to make both small and large changes to their gestural performances through repeated performance. A majority of our participants reported their gestures as becoming simpler and more streamlined, lending credence to our second hypothesis. The rate of redefined gestures leaned heavily towards making more distinct and diverse gestures suggests that users having issues with incorrect recognition resolve the issue redefining the gesture in larger ways in an effort to distinguish it from others. Interestingly, when a gesture was observed as being ambiguous with another, all similar looking gestures were typically redefined to be more distinct rather than just one of them.

The quantitative results did not have as strong a result. Of the three participants that reported producing more complex or elaborate gestures in the questionnaire, none of them reported any of the largest positive variations in distance. Conversely, of the three largest positive distances (participants 5, 8 and 14), only one reported making their gestures more complex, the other two suggesting they were simplified and more compact. This inconsistency in user reporting may be attributed to the user considering the way they defined a gesture and the way they performed it during the trial separately. Greater analysis into user performance during the trial, including the performances of gestures in the experiment compared to the definitions to see if users see them separately.

Future work would expand upon this trial by the inclusion of much greater control over the conditions in future experiments and analysis of user interaction in-trial. An experiment with a control, providing users with a pre-defined set of gestures to perform rather than having their own may also yield interesting results. Future reports may also attempt to report in more detail on how users perceive their gestures as ‘complex’ and simple’ in comparison to others, both mechanically and cognitively.

REFERENCES

Beringer, N. Evoking Gestures in SmartKom - Design of the Graphical User Interface. (2002), 228-240.

Cabral, M. C., Morimoto, C. H., Zuffo, M. K. On the usability of gesture interfaces in virtual reality environments. Conference Name (2005), 100-108.

Cassell, J. A Framework For Gesture Generation and Interpretation. (1998), 191-215.

Dahlbäck, N., Jönsson, A., Ahrenberg, L. Wizard of Oz studies — why and how. Knowledge-Based Systems 6, 4 (1993), 258-266.

Fiedler, A., Gabsdil, M. Supporting Progressive Refinement of Wizard-of-Oz Experiments. Proceedings of ITS - Workshop on Empirical Methods for Tutorial Dialogue Systems (2002), 62-69.

Fothergill, S., Mentis, H., Kohli, P., Nowozin, S. Instructing people for training gestural interactive systems. Conference Name (2012), 1737-1746.

Grandhi, S. A., Joue, G., Mittelberg, I. Understanding naturalness and intuitiveness in gesture production: insights for touchless gestural interfaces. Conference Name (2011), 821-824.

Hincapié-Ramos, J. D., Guo, X., Moghadasian, P., Irani, P. Consumed endurance: a metric to quantify arm fatigue of mid-air interactions. Conference Name (2014), 1063-1072.

Kratz, L., Smith, M., Lee, F. J. Wiizards: 3D gesture recognition for game play input. Conference Name (2007), 209-212.

Malizia, A., Bellucci, A. The artificiality of natural user interfaces. Commun. ACM 55, 3 (2012), 36-38.

Nielsen, M., Störring, M., Moeslund, T., Granum, E. A Procedure for Developing Intuitive and Ergonomic Gesture Interfaces for HCI (2004), 409-420.

Wachs, J. P., Kölsch, M., Stern, H., Edan, Y. Vision-based hand-gesture applications. Commun. ACM 54, 2 (2011), 60-71.

Wobbrock, J. O., Morris, M. R., Wilson, A. D. User-defined gestures for surface computing. Conference Name (2009), 1083-1092.

Wright, M., Lin, C.-J., O'Neill, E., Cosker, D., Johnson, P. 3D Gesture Recognition: An Evaluation of User and System Performance. (2011), 294-313.