EVALUATING TANGIBLE OBJECTS FOR MULTIMODAL INTERACTION DESIGN

Ronnie Taib\textsuperscript{1,2}, Natalie Ruiz\textsuperscript{1,2}

\textsuperscript{1}Interfaces, Machines and Graphic Environments
IMAGEN, National ICT Australia
Eveleigh NSW 1430, Sydney, Australia
{ronnie.taib, natalie.ruiz}@nicta.com.au

\textsuperscript{2}School of Computer Science and Engineering
University of New South Wales
Sydney, Australia

ABSTRACT

The design of applications with multimodal interfaces currently implies complex handcrafting by interface experts, lack of compliance with industry standards of cost effectiveness, maintenance and user focus such as those achieved by the current User-Centered Design methods. This paper presents an initial step towards a design by-example approach, whereby the end-user’s multimodal preferences for a specific domain can be learned during the design phase. In particular, we propose to alleviate the design costs by using tangible objects for designing multimodal user interfaces. Heuristic evaluation shows small to no effect on the user’s preferred multimodal behaviour when comparing tangible and virtual objects during design.

KEYWORDS: Multimodal User Interaction design, User-Centered Design, Design by example.

1. INTRODUCTION

Extensive research on speech and gesture based Multimodal User Interfaces (MMUI) started with Bolt in 1980, later addressing semantic fusion of pen and speech input (Oviatt, DeAngeli, & Kuhn, 1997) and more recently providing untethered gesture recognition (Schapira & Sharma, 2001). However, designing such systems still relies on complex, carefully handcrafted, behaviour rules, such as the various languages and structures required to describe application knowledge, presentation goals and wireframe construction in the WIP planning and presentation system (Wahlster, André, Finkler, Profitlich, & Rist, 1993). More recently, the Embassi system has embedded a complex combination of several ontologies and Description Logics to model the application domain and slightly more generic interaction behaviours (Forkl & Hellenschmidt, 2002). The complexity involved in the design of these output generation modules is equally taxing in the design of input fusion and dialogue management components.

We planned our long-term research along the lines of designing multimodal interaction patterns “by-example”, where machine learning algorithms such as Ripple Down Rules introduced by Compton and Jansen (Compton & Jansen, 1989) are an appealing solution. A design framework can then learn multimodal behaviours by adding exceptions to predefined rules, formalising the reason for the exception e.g. “Incorrect. When I move this way [+gesture], you should rotate forward instead.”

In order to collect preliminary data for the construction of the framework itself, we require a number of interaction sessions for multiple applications. To optimise this process yet following User-Centered Design (UCD) for formative experiments, we have to address the costs and appropriateness of some classical methods. Despite an interdisciplinary overlap, efforts to apply UCD in the MMUI domain have often been limited to high-level guidelines such as measuring the adequacy of expected, sensed and desired actions, as proposed by Benford et al. (Benford et al., 2005). In our experience, some classical methods have proven inappropriate: paper mock-ups are cumbersome for multimodal input and output, and electronic mock-ups are often too costly to implement when several modalities are involved. The
Wizard of Oz (WoZ) approach is popular in multimodal research, for evaluative or formative purposes, as used by Hauptmann to carry out a study on user preferences in terms of speech and hand gesture interaction with a 3D cube (Hauptmann, 1989). Although it may be beneficial to collect subject reactions to a particular interface, the WoZ approach does not encourage subjects to teach the system how to improve or customise its operations as is required by reinforced learning methods. In this paper, we describe an hybrid approach, whereby the experimenter plays the role of a system agent in order to learn preferred usage patterns from end-users. In particular, we explore the feasibility of using tangible objects manipulated by an agent as an alternative to building an expensive electronic mock-up. The aim is to assess whether subjects would interact similarly with a tangible object and with its virtual representation. The next section describes the user experiments carried out; followed by a presentation and discussion of the results. The paper concludes on the potential benefits of this approach and describes some future work.

2. USER EXPERIMENTS

2.1. 3D-Graph Navigation Application

The first application chosen for the experiment was a 3D visualisation tool where users can manipulate 3D-graph structures (see Figure 1), essentially translating and/or rotating them in space using speech, head and hand gestures. To our knowledge it is a novel interface for this type of tool. Hence, the experiment has a formative aspect beyond the scope of this paper; it is similar to (Hauptmann, 1989), with the addition of head gesture, proposing to determine preferred multimodal interaction. The first motivation though was to collect patterns of design by example for multimodal interaction as well as comparing tangible and virtual representations during this process. The experimenter played the role of a learning system agent (no human-human interaction was allowed) responsible for moving a tangible or virtual object according to the subject’s instructions.

2.2. Methodology

The experiment involved 16 unpaid subjects, 8 females and 8 males, aged 18 to 50. Most subjects had no prior knowledge at all about 3D-graph visualisation and navigation.

We introduced 8 complex conditions: combining 2 “representation” conditions which included a tangible object and a virtual version of the same object as shown in Figure 1; and 4 “interaction” conditions including speech only, hand(s) gesture only, head gesture only and the combined used of those 3 modalities (multimodal condition). The order in which the representation and interaction conditions were presented was randomised independently between subjects.

Figure 1: Subject Interacting with the Tangible Object (left) and Virtual Representation (right)

Each subject was asked to perform a similar task under each of the 8 conditions assigned to him, which took between 20 and 25 minutes for all conditions. The main task was to navigate the simple 3D graph shown in Figure 1 given specific start and end positions. The subjects were asked to perform the task using 3D movements along one major axis at a time (e.g. rotation around Z axis), and were encouraged to employ as many moves as possible in order to perform this task.
The experimenter acted as an agent moving the tangible object in space or the virtual representation in a 3D visualisation software according to his interpretation of the subject’s multimodal instructions, initially using an arbitrary mapping between subject movements and required actions. The subjects are not allowed to move the object by themselves. They were instructed never to adapt to the ‘agent’ and to notify any incorrect interpretation made by the agent. In such cases, the object was replaced at the previous position, the subject would provide the same instruction again and an alternative, equally viable, interpretation was made by the ‘agent’. Should a second consecutive failure occur, the subject then had to explain the expected interpretation (using plain English or eventually moving the object themselves).

2.3. Experimental Set-up

The subjects were standing at 2 metres of either the ‘agent’ (the experimenter) or of a large screen (2x1.5 m). All interactions were captured by a front camcorder and a side web cam, located at about 2m each. The users were allowed to move slightly around their mark if needed. A paper questionnaire was completed, inquiring about the preferred modalities in terms of how intuitive, comfortable, effective and pleasant they were as well as how different the subjects perceived their interactions with a tangible vs. virtual object.

3. RESULTS AND DISCUSSION

3.1. Tangible vs. Virtual Interaction

Subjects used a 9-point Likert scale to rank how differently they perceived their interactions to be when using the tangible object vs. the virtual representation of that object, 1 being absolutely identical and 9 for an extremely different behaviour. The average rating for the group is 2.8 (standard deviation $\sigma=1.8$) showing that the subjects felt only a very mild difference between the two presentation modes.

We also analysed the discrepancy observed between subjects exposed to the tangible then virtual (T-V) presentation conditions vs. virtual then tangible (V-T) conditions. It appears that T-V subjects reported 3.6 on the same scale ($\sigma=2.1$) whereas (V-T) reported 2.0 ($\sigma=0.9$). However, a one-way ANOVA analysis showed those results are not significantly different from one another ($p=0.0682$).

Even though more results may be required to confirm this finding, it seems that on a subjective viewpoint, the impact of tangible vs. virtual representation does not affect the way subjects perform the task, making tangible objects good candidates for a multimodal design framework. The most obvious advantages we foresee are the cost reduction and the possibility to gather data for reinforced learning.

3.2. Modality Evaluation

The average modality ranking over both types of representation over intuitiveness, comfort, effectiveness and pleasantness, is shown in Figure 2 (1: easiest, 9: hardest) and seems to suggest the superiority of multimodal interaction over unimodal interaction in our sample group. The hand gesture modality is ranked a very close second across all dimensions, probably due to the tangible characteristics of the data. Surprisingly, speech only input comes a close third, despite the spatial manipulation task. It is found to be less effective than hand gesture but approximately as pleasant. These observations cast a different light on the findings of Hauptmann (Hauptmann, 1989) who showed that multimodal interaction over unimodal interaction in our sample group. The hand gesture modality is ranked a very close second across all dimensions, probably due to the tangible characteristics of the data. Surprisingly, speech only input comes a close third, despite the spatial manipulation task. It is found to be less effective than hand gesture but approximately as pleasant. These observations cast a different light on the findings of Hauptmann (Hauptmann, 1989) who showed that multimodal is relatively preferred over unimodal, but this preference may not be as dramatic on an absolute scale, as shown here. A one-way ANOVA reveals the modalities are significantly different on all dimensions ($p<0.001$), namely the head-gesture modality is significantly less intuitive, comfortable, efficient and pleasant – overall least preferred. This reflects the difficulty subjects experienced in finding an appropriate mapping between gestures and 3D navigation.

The use of the experimenter-agent described here can be cost-effective in more general formative studies, making it a potential candidate for multimodal UCD process. However, usability studies should be carried where a predefined mapping between modalities and object movements is proposed to the subjects, in order to validate whether the modality itself is completely inappropriate or whether coming up with a mapping on-the-fly is the main hurdle.
4. CONCLUSION

Some classical User-Centered Design methods such as paper or electronic mock-up are not well appropriate for addressing multimodal interaction, hence have not been applied widely in that community. This paper presented an evaluation of the method of using tangible objects as multimodal interaction design tool in order to focus on the end-user in a cost-effective way. The evaluation results are encouraging, showing no perceived discrepancy in interaction mode between a tangible and virtual representation of the domain data. The method did not impact the overall modality comparison which also carried out during the experiment. Further experiments and analysis are required to determine the applicability of the method described, especially with more complex and non-spatial tasks, but we believe it should be applicable to less tangible type of data (e.g. email application).

Using such methods should help us streamline interaction sessions across a range of applications where the subjects teach the system how to refine default multimodal input interpretation, eventually leading to building an automated design system able to acquire multimodal behaviours by example.

5. REFERENCES


6. ACKNOWLEDGEMENTS

The authors would like to thank the experiment volunteers for participating in the data collection.