MozArt: A Multimodal Interface for Conceptual 3D Modeling

Anirudh Sharma, Sriganesh Madhvanath, Ankit Shekhawat, Mark Billinghurst*
Hewlett-Packard Labs, Bangalore, India
*HIT Lab NZ, University of Canterbury, NZ

{anirudh.sharma, srig, ankit.shekhawat}@hp.com, mark.billinghurst@hitlabnz.org

ABSTRACT

There is a need for computer aided design tools that support rapid conceptual level design. In this paper we explore and evaluate how intuitive speech and multitouch input can be combined in a multimodal interface for conceptual 3D modeling. Our system, MozArt, is based on a user's innate abilities - speaking and touching, and has a toolbar/button-less interface for creating and interacting with computer graphics models. We briefly cover the hardware and software technology behind MozArt, and present a pilot study comparing our multimodal system with a conventional multitouch modeling interface with first time CAD users. While a larger study is required to obtain statistically significant comparison regarding efficiency and accuracy of the two interfaces, a majority of the participants preferred the multimodal interface over the multitouch. We summarize lessons learned and discuss directions for future research.

Categories and Subject Descriptors

H.5.2 **[Information Interfaces and Presentation]:** User Interfaces – *Input devices and strategies*

General Terms

Design, Human Factors

Keywords

Multimodal interfaces, CAD, 3D modeling, Speech input

1. INTRODUCTION

Computers have become more affordable and accessible, but 3D modeling remains a complex task involving a steep learning curve and extensive training. There are several reasons for this, such as the need to learn a dense toolbar and menu based WIMP interface, and use of effectively only one I/O device (either the keyboard, or the mouse)[10]. This increases the application learning time.

In our research we are interested in the problem of designing a 3D modeling application for conceptual modeling. During early stage concept design the user's attention is focused on overall appearance of the model, and exact dimensions, positions and tolerances are dealt with during the later design phase. Unfortunately, most CAD interfaces force the designer to define the model precisely through a large number of icons and menu options even at the concept modeling stage. Another key issue is the extensive use of keyboard and mouse input to operate a complex GUI. It takes a substantial effort to attain design

proficiency with these tools, and it is commonly observed that designers prefer pen and paper for early stages of design.

In this paper, we present *MozArt*, a prototype interface that explores how multimodal input and appropriate hardware may be applied to simplify conceptual 3D modeling for first-time CAD users. MozArt features a minimalist UI, and uses a touch table with tiltable orientation. Generally speaking, a drafting table is a preferred means of drawing, commonly used by artists and architects during drawing and modeling tasks. Studies show that a tiltable touchscreen is more efficient than its vertical counterpart since it allows resting of elbows upon the bezel of the screen [3].

This paper is organized as follows. We first cover related work in multimodal interfaces for computer modeling. Next, we describe the MozArt hardware prototype - a tilt-able drafting-table style interactive touch surface. In Section 4, we describe the MozArt user interface, and the touch and speech interactions we have enabled for 3D modeling. Next, we present results from an initial user study of user performance using multimodal and pure multitouch interfaces. We conclude the paper with a discussion of lessons learned and directions for future research.

2. RELATED WORK

Beginning with Bolt's *"Put that there"* [15], there have been many interfaces that show the value of combining speech and gesture input for graphical applications. Boeing's *"Talk and Draw"* [16] application allowed users to draw with a mouse and use speech input to change UI modes, and was one of the first multimodal drawing applications. In [9], speech and glove based gesture input were integrated with a stand-alone CAD package. Multimodal interaction has also been used in 3D graphic environments and compared to traditional interfaces; e.g. Hauptmann [1] investigated the use of multimodal interaction for a simple 3D cube manipulation task and found that people strongly preferred using combined gesture and speech for the graphics manipulation, instead of either modality alone.

These systems typically used pen or glove based input devices, but recently multitouch table-top displays have made it possible to directly manipulate 3D graphics. For example, SpaceClaim demonstrated a prototype CAD tool with a multitouch screen [17]. However this application does not easily support task-switching and continues to use the same toolbars and buttons as used with a keyboard and mouse metaphor. The work of Arangarasan and Gadh [2] makes use of multimodal input and output for working with 3D models, however, it requires the use of a Head-Mounted Displays (HMDs) that can be obtrusive to beginners due to ergonomic issues. ILoveSketch [4] combines stylus input with automation to allow users to create accurate 3D models primarily by sketching using an electronic stylus. SenStylus [11] demonstrates a stylus based 6DOF input system for 3-D modeling. While well suited for the creation of curved and organic shapes, these interfaces may be more complex than necessary for simple

3D design. Computer modeling has traditionally been viewed as the domain of experts, and these systems are in general targeted towards their needs.

In contrast to these efforts, our emphasis is on supporting 3D modeling by novice users using a minimalist multimodal interface on appropriate hardware, and attempting to ensure that the user's focus is on the design task rather than the UI elements.

3. MOZART TABLE

In this section, we describe the MozArt table, the hardware prototype that we have built. The prototype draws inspiration from the architect's drafting table. It is a common observation that designers often work on their sketches and design drafts on a table that allows them to change orientation of the base plate, while resting their hands and elbows on the bezel. This affords idea sharing by virtue of the table's size, visibility, and provides a platform that enables discussion about designs [5]. The MozArt table involves a fabricated tilt mount with a multi-touch display with which the user can interact using voice and touch (Figure 1).

Figure 1: MozArt Table hardware prototype

Multi-touch detection is implemented using the optical Laser Line Plane technique [14]. The surface is illuminated with 850nm Infrared lasers. Scattered light generated upon touching the surface is captured by a modified infrared camera below the table surface, and is processed using Community Core Vision libraries [8] to detect and track touchpoints. The system is capable of tracking up to 255 touch-points with an accuracy of 3mm.

4. MOZART USER INTERFACE

Our chief goal was to explore a natural and minimalist multimodal interface for 3D modeling allowing the user to focus on the design process rather than the user interface components. To create a multimodal design tool, we decided to add support for speech and gesture input to an existing modeling tool, SketchUp [12]. It is worth noting that by default, the SketchUp UI is designed for keyboard and mouse interaction, and features extensive use of menus for selecting modes, views and shapes. For our prototype, we designed a custom interface, which replaced the use of toolbars and buttons for a subset of common tasks, with touch and speech commands.
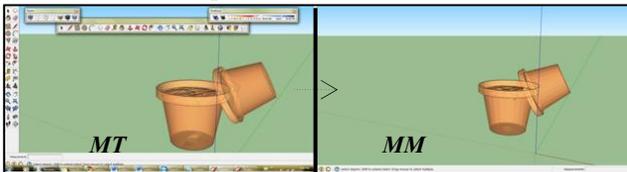

Figure 2: Default SketchUp UI transformed into minimal UI

The vocabulary of Touch and Speech commands is shown in Table 1 and Table 2 respectively.

Table 1: Touch Gestures

Touch Gesture	Outcome
Two finger pinch	Zoom in/ Zoom out
Single touch drag	Specify extent of M and O

Table2: Speech Commands

Speech input	Category	Mode
"Circle"	Activate (circle)	Modify (M)
"Rectangle"	Activate(Rectangle)	Modify (M)
"Line"	Active(Line)	Modify (M)
"Pull it up"	Activate(Push/Pull)	Operate (O)
"Show Isometric"	Show isometric model view	Change viewport(V)
"Top View"	Show top view	Change viewport(V)
"Show me around"	Activate 3D orbiting	Change viewport manually (V)
"Undo"	Undo last step	Modify (M)

Figure 3 shows the drawing of a square, extruding it into a cube, and zooming in and out using speech and multitouch gestures.
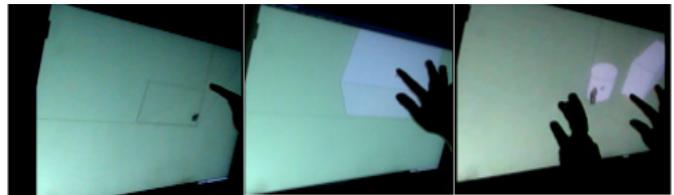

Figure 3: Draw/Extrude/Zoom modes of the multimodal UI
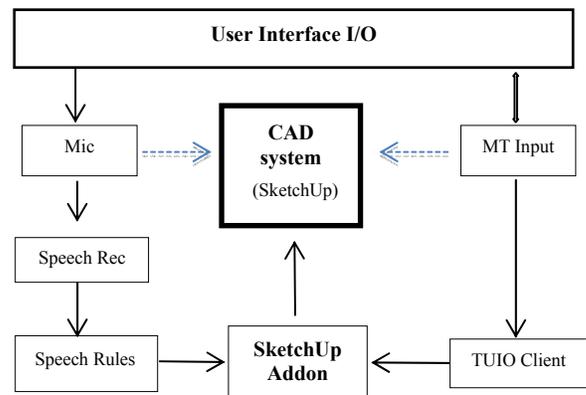
4.1 System Design

Figure 4: System Design

The multitouch data from the table is captured in TUIO[13] format, parsed using a client written in C++ and loaded into SketchUp using a Ruby script. The multitouch gestures are then mapped to events such as dragging, zooming, and panning. CMUSphinx [7] is used to process speech from a microphone. The commands are forwarded to SketchUp to initiate keyboard shortcuts for switching between tasks (Figure 4).

4.2 User Evaluation

MozArt is intended to simplify the design process for users familiar with computers but not expert at 3D CAD modeling. In order to evaluate MozArt, we conducted a user study with 8 male and 4 female participants from an office environment. The participants were 20-29 years old. The user evaluation consisted of a training phase followed by a testing phase.

Training Phase

Each subject was shown a 5-minute introductory video about face and line modeling concepts using SketchUp's conventional mouse/keyboard interface. They were then allowed to try the multitouch gestures and menus for 5 minutes. Being first time CAD users, most of the subjects tried drawing, and extruding objects, similar to what was shown in the introductory video.

Testing Phase

The testing phase consisted of subjects completing two modeling tasks of different complexity. For each task, subjects were shown one model on-screen and were asked to make a copy of it. The models were designed to test basic modeling skills such as drawing polygon faces, extrusion, navigating the UI, and so on. Each modeling task was completed by the user using two interface conditions, Multitouch and Multimodal.

Multitouch only (MT)

For this interface condition, a conventional Toolbar/Button based touch interface was enabled. The interaction tools had to be selected from the interface menu. The size of the buttons was increased to reduce touch-based selection errors. The user was constrained to use only multitouch gestures.

Multimodal (MM)

For this condition, the Toolbars/Buttons were disabled and hidden. Only the blank modeling canvas was presented to the user and speech and touch based interaction was enabled. The user was constrained to use speech commands together with multitouch gestures. A printed sheet listing the available speech commands was provided to the user.

One half of the participants used MT followed by MM, while the other half used MM first. As participants completed each modeling task the following experimental measures were collected:

- Task completion times
- X-Y coordinates of touch-points on the screen
- User errors measured as the number of "undo" commands
- NASA Task Load Index surveying user workload

At the conclusion of both tasks using both techniques, participants were asked to state and discuss their personal preference of technique.

Average Task Completion Times

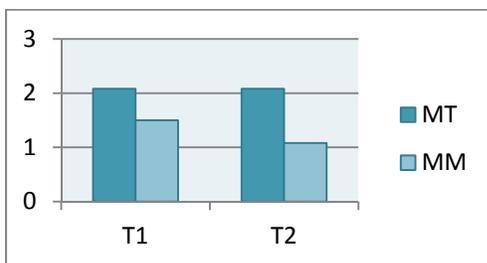The average task completion times (in minutes) for different tasks and techniques are shown in Figure 5.

Figure 5: Average task completion times (in minutes)

We compared task completion times using a two factor ANOVA test with replication alpha=0.05, which yielded F= 0.468 < $F_{critical}$= 4.062, p>0.05. This shows that the influence of the technique on the task completion time is not significant. However, we did observe that some of the task completion times were unusually large. For example, one of the participants with unusually large hands had issues selecting the correct menu buttons in the MT condition. Another user had an accent that was not recognized well by the speech recognizer.

Errors

The number of errors was defined as the number of times a user mis-selected a tool and used Undo. This was measured for each condition. Figure 6 shows the average number of errors per user for each task and condition. We used a one-tailed t-test to test whether the MM condition produced fewer errors for Task 1 than the MT condition, and found (t(11) = 1.02, p = 0.164). For Task2 a one t-tailed t-test found (t(11) = 3.07, p = 0.005). Thus for Task 2 the MM condition produced significantly fewer errors than the MT condition, whereas for Task1 there was no significant difference in average number of errors.
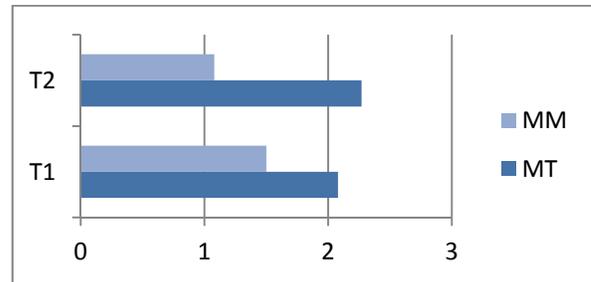

Figure 6: Average number of errors per user

The decrease in number of errors in the MM method for Task 2 compared to Task 1 is potentially due to the user's increased familiarity with speech commands. Despite designing a more complex model during Task 2 participants committed fewer errors since they had become familiar with the simple speech commands and hence did not have to refer to the printed sheet of commands as frequently.

On the contrary, in the MT method we noticed that as the complexity of the model increased from Task 1 to Task 2, the participants had to move their hands, resulting in a greater number of touch based false positives.

Heat maps

A visual comparison between multitouch and multimodal conditions was enabled using heatmaps plotted from the complete set of logged touch coordinates (Figure 7). Based on these heatmaps and our discussion with the participants, it was evident that the MT condition required participants to shift their visual attention more often between model and toolbars/menus at the top and sides of the screen. It was observed that every time the user changed a tool, their hand and focus of attention needed to shift to the top and sides of the screen and back to the center. However, for the MM condition, touch interaction took place primarily on the model itself.
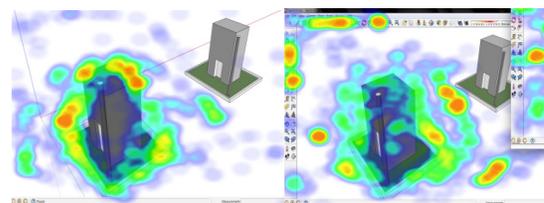

Figure 7: Heatmap generated from touch coordinates for MM (left) and MT (right) . Darker color implies more points.

Task Load Index

We used the NASA TLX method to collect subjective task load data after each modeling task. Upon analysis we found a general

trend towards increased frustration and physical demand for the MT condition (Figures 8 and 9).
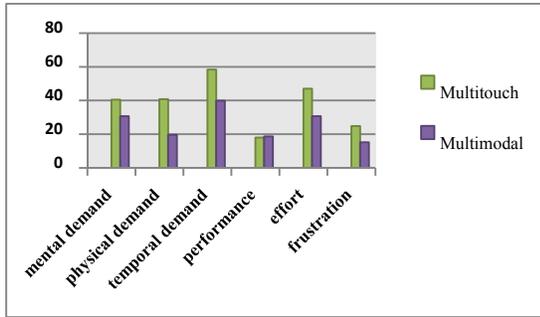

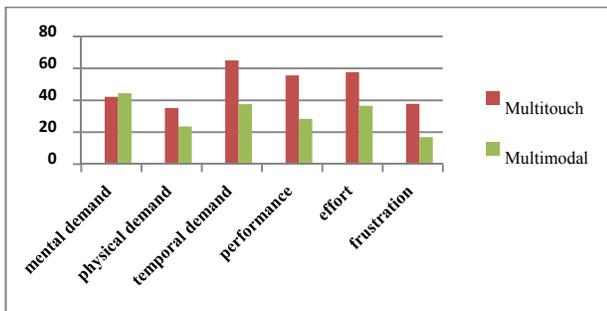**Figure 8: Task Load Index for Task1**

Figure 9: Task Load Index for Task2

The higher frustration may be attributed to the higher frequency of errors, and increased physical demand to the fact that considerable hand movement was required every time a tool had to be changed.

Subjective Preference
On completion of both tasks, all participants were given an opportunity to select one of the two interfaces and build a model of their own choice; 9 out of 12 participants chose the MM interface while 3 chose the MT interface. Participants also reported on difficulties faced with the MT interface, e.g. ones with sweaty and/or large fingers reported problems selecting the exact buttons from the toolbar, leading to larger task completion times.

6. SUMMARY AND FUTURE WORK
We have described our prototype MozArt system, a tiltable multi-touch table with a multimodal interface that attempts to present an easier way to create 3D models. The form factor chosen provides a greater sense of directness because participants put their fingers directly onto the item of interest and manipulate it. The tilt enhances collaboration and prevents fatigue over prolonged working sessions [3]. MozArt allows the use of speech commands to perform actions that would ordinarily require the manipulation of menus and icons. By so doing, we believe the interface allows the user to focus on the task rather than the interface itself.

We evaluated the system with a user study. We conducted a Two Factor ANOVA with replication and found the results to be statistically insignificant, primarily due to the small number of participants taken for the study. Other factors were the presence of outliers in some of the methods. Our next step is to repeat the study with a larger number of participants so that any differences in task efficiency and accuracy between the use of multitouch only and multimodal techniques may be brought out. We would also like to include a Likert test to study subjective preference for the two techniques.

In the future we will work on further expanding the multimodal vocabulary. For many modeling tasks, precision of extents becomes a key issue. To address this, we intend to support spoken commands that allow precise dimensions to be specified.

Finally, speech input using an open microphone suffers from the problem of false positives (spurious input) due to ambient sounds, especially in multi-user environments. This could be solved by adding lip movement detection based on image processing to detect when the user is speaking.

5. ACKNOWLEDGMENTS

6. REFERENCES
[1] Alexander. G. Hauptmann: "Speech and Gestures for Graphic Image Manipulation." CHI 1989 241-245.

[2] Arangarasan, R., and Gadh, R., "Geometric Modeling and Collaborative Design in a Multi-modal Multi-sensory Virtual Environment." ASME DETC'00

[3] A. Sears. "Improving touchscreen keyboards: design issues and a comparison with other devices." Interacting with Computers, 3(3), 1991.

[4] Bae, S.-H., Balakrishnan, R., And Singh, K.. "ILoveSketch: As-Natural-As-Possible Sketching System for Creating 3D Curve Models." UIST, 2008.

[5] Buxton W., Fitzmaurice G., Balakrishnan R., Kurtenbach G., 2000, "Large Displays in Automotive Design", IEEE Computer Society 20 (4), 68–7.

[6] C. Muller-Tomfelde, A. Wessels, and C. Schremmer, "Tilted Tabletops: In Between Horizontal and Vertical Workspaces". TABLETOP '08.

[7] CMUSphinx Libray http://cmusphinx.sourceforge.net/

[8] Community Core Vision http://nuicode.com/projects/tbeta

[9] D. Weimer and S. K. Ganapathy, "A synthetic visual environment with hand gesturing and voice input," CHI 1989, pp. 235–240.

[10] Darren A., Yaser G., Frank Maurer 2010 "Adapting existing applications to support new interaction technologies: technical and usability issues". ACM SIGCHI symposium on Engineering Interactive Computing Systems.

[11] Fiorentino, M., G. Monno And A. Uva . "The Senstylus: a Novel Rumble-Feedback Pen Device for CAD Application in Virtual Reality", WSCG 2005.

[12] Google SketchUp http://sketchup.google.com/

[13] M. Kaltenbrunner, et. al, "TUIO - A Protocol for Table Based Tangible User Interfaces," in GW '05: Proc. of the 6th International Workshop on Gesture in HCI and Simulation, 2005.

[14] Park, J. and Han T.: LLP+: multi-touch sensing using cross plane infrared laser light for interactive based displays, SIGGRAPH Posters, 2010.

[15] Richard A. Bolt. 1980. "Put-that-there: Voice and gesture at the graphics interface", SIGGRAPH 1980, 262-270.

[16] M. Salisbury , J. Hendrickson , T. Lammers , C. Fu , S. Moody, Talk and Draw: Bundling Speech and Graphics, IEEE Computer, v.23 n.8, p.59-65, August 1990.

[17] SpaceClaim CAD http://www.spaceclaim.com/en