School of Computing

University of Teesside

Middlesbrough

TS1 3BA

Human Computer Interaction within Industry Tools

Bsc. Computer Games Programming

Sam Oates

Supervisor: Tyrone Davison

Second Reader: Suiping Zhou

**Abstract**

This paper follows the development of a 3D computer games tool powered by a human computer interaction based device, the Microsoft Kinect.

Research was based around three fundamental areas required for the project. Human computer interaction (HCI), real-time image recognition and the deformation of terrain within 3D graphics.

Using previously gained industry knowledge and details gained from my areas of research, an initial design prototype was created, followed by a small amount of user testing. Testing for ease of use, productivity and comparing against gestures natural within the real world.

**Acknowledgements**

I would like to thank my two managers whilst working at Blitz Games Studios. Neil Holmes and Tom Gaulton, both of which encouraged my passion towards computer games tool systems. I would especially like to thank Neil Holmes for giving thorough feedback on the project in its later stages of development. Along with this Terry Greer, a designer a met whilst working at Blitz Games Studios helped out when drafting up initial concepts and design ideas. Finally I would like to thank every person whom tested the project, both at the designated times and at random points during the development cycle.

**Contents**

**Abstract**

**Acknowledgements**

1. **Introduction**
   1. Research Question
   2. Rationale for Project Choice
   3. The Current State of Human Computer Interaction
   4. System Requirements
2. **Methodology**
   1. The Methodology behind the Implementation
   2. Design Methodology
   3. Methodology of Testing
3. **Research Analysis**
   1. Real-Time Hand Detection
   2. Voice Recognition and Synthesization
4. **Design and Implementation -> C++, SDKs, UMLs**
   1. Programming Languages and SDKs
   2. Project Design
5. **Testing: Round One**
   1. A
   2. B
   3. C
6. **Reevaluating the Design and Implementation**
   1. A
   2. B
7. **Testing: Round Two**
   1. A
   2. B
8. **Evaluation**
   1. A
   2. B
9. **Conclusion**
10. **References**
11. **Bibliography**
12. **Appendices**

## Introduction

### Research Question

In the modern day games studio, artists and designers are often found using keyboard and mouse input to create scenes, art assets and such; for games. However, creative people have a tendency to work better with their hands. The keyboard and mouse input may limit their ability to do this. Posing the question, is current computer hardware limiting usability with its non-natural interface?

I aim to create a simple tool (in the form of a terrain editing system), where the input is based upon the user within their 3D environment (via the use of the Microsoft Kinect device) as well as using other inputs such as the users’ voice. This creates an interface more in tune with its users’ tendencies resulting in the exploration of the users’ potential productivity gain and a potential higher quality of work. Where by the main complication in implementation will be finger tracking and hand gesture recognition, due to variations in hand size and shape of different users as well as different mentalities of how they believe the gestures should work.

### Rationale for Project Choice

I have had a life-long passion for tools within computer games, trying to make interfaces and systems as simple as possible for the user to interact with. My inspiration for this project was found whilst on work placement at ‘Blitz Games Studios’. Whilst there I spent time working on their tool system (‘Blitz Tech’) as well as working closely with game teams and at points the Microsoft Kinect. Whilst working I noticed how the artists, designers and animators used real-life models and scenarios to compare with their plans or creations. Using pen and paper as well as other input devices such as tablets to draft work before creating the asset within a 2D or 3D graphics computer tool.

With this, I have first-hand experience of how an artist works and how a programmer creates software. However the two do not necessarily correlate due to the differences in rational between artists and programmers. To expand on this, I have experience with user interfaces, tools graphics/rendering and the Microsoft Kinect.

### The Current State of Human Computer Interaction

Human computer interaction (HCI) is an astronomical field of ongoing research. However the majority of such research is specific towards the general user and or non-computer user, attempting to allow non-technical people to interact with computer hardware. The problem lies in extracting data from the user in a manor most natural to them and evaluating the data for use with a device, as doing such is hard to generalize. This results in software that feels natural to some and not to others.

This problem is reduced when looking into to HCI within the games industry, as we can make the assumption that the user is somewhat technically minded. Already the user should have an understanding of current HCI making use of the standard keyboard and mouse, as well as other artist specific input devices.

### System Requirements

Given the problem of non-natural HCI interfaces for creative peoples, specifications for a natural HCI interface can be formed.

An artist should be able to move a gizmo (a replacement for the mouse cursor in three dimensional space) about the terrain environment using nothing but there hand. Once positioned to the users requirements, the user need only use their other hand to apply the selected brush to the terrain in an area about the gizmo.

The user should be able to change the selected brush via a graphical user interface (GUI) based menu system, via the use of voice commands and hand gestures. The GUI menu system should also allow access to other mandatory tasks associated with a terrain based tool system. This includes creating a new default terrain, opening existing terrains and saving the currently active terrain. Along with this editing settings about the size and strength of the currently active brush should be performed via voice and gesture based commands however this does not require the user to traverse a menu system and instead should be performed at any point during runtime.

To achieve the above outcomes the following steps need to be fulfilled.

* Implement a simple C++ terrain rendering system using Microsoft’s Direct3D 11.
* Allow the terrain system to be deformed via the use of the traditional keyboard and mouse.
* Consult potential users on gestures for different operations.
* Implement the first draft of gesture based commands.
* Test first implementation with potential users.
* Improve first draft based on feedback from initial testing
* Consult potential users on voice commands for performing different operations within the system.
* Implement first draft of voice based commands
* Test improved gesture commands and draft voice commands with potential users.
* Finalize gestures and voice commands based upon user feedback collected during testing.

Overall, the goals for the project are as follows.

* Create a simple terrain editing system powered by a natural human based input device.
* Attempting to prove that both productivity and quality can be improved by reducing the barrier that exists between creative people and the tools extant within industry.

## Methodology

### The Methodology behind the Implementation

As the product requires feedback based upon user experience, the product has to go through multiple repeated steps of development until all discovered issues are resolved. This means should there be an unseen problem within the initial plans; it can be refactored out at a later stage of development. The Kinect device runs at a low resolution meaning sampling hand data has potential issues. The recursive development cycle can help resolve the issues should initial implementations be unsuccessful. Three public testing points are to be set, where artists, designers and other creative will be invited to try out the project in its current development state. After each test session feedback will be collected and development tasks reassessed.

Due to the rapid changes that will take place based upon user feedback; a form of source code versioning control software is required. Whilst working in the industry we used subversion control (SVN) [1]. SVN is based around the principle of one main repository (the server) and multiple local copies (the client(s)). A client need simply check out the latest revision from the server to create a local copy. Changes are then made to the local copy and committed to the server. Each revision stores additions, deletions and changes to source files. This allows the client to revert back to a previous version of the code base, if required.

Along with this, SVN has the ability to branch and tag revisions. A branch allows a developer to work in parallel to the main repository. Commits made by the developer are made into their branch of code rather than the main repository allowing large scale features to be implemented without breaking the main repository. Once the feature has been completed the branch is merged back into the main repository. Tagging can be used to flag a given revision. For example the version of the project used for each test process will be tagged as such. Meaning in the future returning to specific versions can be performed to compare and contrast both the source code and the features.

Figure 1: http://www.hosting.com/media/387108/svndiagram\_500x416.jpg

With each commit a comment can be entered detailing important changes and additions that occur within the commit, making finding older revisions easier to pin point. The ability to revert to previous versions of the project will help in finding and resolving many issues that may occur during development.

### Design Methodology

Given that hardware tessellation is implemented within the Direct3D 11 SDK this will most likely be chosen to be the graphical API, as tessellation can be used to smooth the terrain with little overhead in performance. Due to this implementation shall be done using the program using the C++ programming language. Not only for the easier implementation of hardware tessellation but also for the speed and performance gain which are present with the C++ programming language. The Kinect SDK (which I have previously used) also has a C++ implementation which again has performance gains over other language implementations. The speed benefits from using a low level language such as C++ allows for fast processing of the vast amounts of data that will be gathered by the Kinect camera and its microphones.

The project will be initially drafted via the use of unified modeling language (UML) diagrams. The object-orientated design of the C++ programming language allows UML to easily layout and design classes and interfaces required for the project.

### Methodology of Testing

Throughout the project not only will I test usability and features, but static analysis will be performed on the code before every commit.

#### Static Analysis

Static analysis is the term used for the process of programmatically scanning source code for potential issues. To perform this task an SVN commit hook can be created to perform static analysis via a program called Cppcheck [2]. This means, that prior to any SVN commit static analysis will be run on the added and or changed code. Cppcheck performs the following checks;

* Out of bounds checking
* Check the code for each class
* Checking exception safety
* Memory leaks checking
* Warn if obsolete functions are used
* Check for invalid usage of STL
* Check for uninitialized variables and unused functions

Should there be an issue with any code a report is presented to the user and the commit is cancelled until all static analysis tests pass successfully. This will help vastly in improving the stability of the project as well as pointing out potential mistakes in logic which would previously go unseen.

#### User Based Testing

User based testing will be performed in two parts; three main testing phases and continuous testing with non-project specific users. The continuous testing will be performed whilst developing the project in the universities computer laboratories. The basic principle will be based around people’s interest in the project. Given the interactivity and abstract nature of the project due to the use of the Microsoft Kinect device other people about the computer lab may be willing to test new in-development features at arbitrary points. This will help fine tune features as well as spot issues within the design at an earlier stage of development.

Finally the three designated user test points will be used to test the current state of the project, upon the users the tool is designed for. Each test phase is to be designed to test a specific feature of the tool system. The phases are as follows;

* Basic hand gesture detection and the basics of the terrain system itself.
* Finalizing hand gesture detection, introducing voice based commands.
* Final testing of fully implemented hand gesture and voice recognition commands.

After each session feedback will be taken both verbally and in the form of a short questionnaire. The questionnaire will pinpoint areas which are new to the current test state of the project. The interesting point with the test is that, in every session the user can not only test the Kinect based input, but also all features will be implemented via the keyboard and mouse. This allows the user to properly compare the two interfaces upon the same terrain editing system.

Along with this, in the later stages of the project I shall contact my manager from Blitz Games Studios, Neil Holmes to ask for his professional opinion about the product. With the aim that the project may be passed around the office to some other professionals whom not only work on computer game tools but also artists whom use the tools themselves. Giving a real insight to whether the industry actually believes natural HCI is a possibility within industry level tool systems.

## Research Analysis

### Real-Time Hand Detection

The project lies around the principle of using the Microsoft Kinect as an input device, using nothing more than the users’ hand(s). Although I have previous experience with the Kinect device, hand detection lies out of the bounds of the normal Kinect and its SDK. The Kinect SDK implements full body (skeletal) tracking, but nothing for individual limbs of the human body. It also provides no built in image/object detection methods.

The Kinect device offers two types of data; a depth image and a colour image. Mathew Tang proposes the use of both sets of data to best estimate the hands existence and gesture (or shape) [3]. His technique involves cleaning up the RGB (red, green and blue colour) image using standard image processing techniques like colour balancing and dilation/erosion, then incorporating the depth data using a simple probabilistic model. This was followed by normalizing the rotation of the estimated results. Finally three sets of features are then extracted from the data; the raw pixel estimates, a radial histogram and a modified version of the SURF (speeded up robust features [4]) descriptors. Tang had some success with his approach, though accuracy lied around the 90% threshold. Tangs technique also had the limitation of only detecting a single hand, where as I aim to support multiple hands.

Du and To [5] suggest using a different Kinect SDK from the official Microsoft SDK. They suggest using OpenNI (Open Natural Interaction), an open source alternative. However there implementation follows the same principle as Tang; however they only filter down the hand data based upon the depth data the Kinect produces. This is proceeded by six steps, which are used to calculate the contour of the hand.

Firstly the image is filtered, via the means of a median filter with a sample area of 15x15 pixels. This smooth’s the raw extract depth image. The second stage is to trace the edges of the filtered image, giving the rough contours of the hand. This simplifies the complexity of the rest of the stages. Given the outline of the hand, an estimated polygon can be formed, estimating a low poly shape of the hand. Using an image library such as OpenCV (Open Source Computer Vision Library) this can be done using an algorithm such as the Douglas-Peucker graph algorithm [6]. Given the low poly estimate of the hand and the contours of the hand, stage four is to detect the concavities and the convex points on the hand. Again this task can be performed via the use of OpenCV commands. Once points have been generated upon the contours of the hand, the convex and concave points are filtered. The filter is performed in two passes. Firstly clusters of points where merges into a single point, should the distance between the points fall below a given threshold. Secondly any convex points which fell below the palm of the hand where removed. Finally the resultant points can be used to estimate the number of digits visible on the hand. Du and To’s technique is again limited to the use of only one hand, however has a 94% accuracy rate and a low performance impact.

Figure 2: Starting from the top left; Original depth map, Rescaled image, Background elimination, Extraction

### Voice Recognition and Synthesis

The Kinect hardware contains four microphones in an array. Along with the Microsoft Kinect SDK has support and example projects showing how to use and implement voice recognition. However, although the Kinects array of directional microphones work well, I predict that in the hectic work place background noise may be an issue whilst performing voice commands. A solution to this would be for the user to wear a headset (with microphone). This would mean that background noise would be brought to a minimum, resulting in less false positives of keyword recognition.

Figure 3: The Kinects Microphone Array highlighted in Purple. http://social.microsoft.com/Forums/getfile/18629/

Microsoft offers the Speech API (SAPI), which supports speech recognition and speech synthesis (via the Microsoft text-to-speech (TTS) engine). However research suggests that the official Kinect SDK implements part of SAPI internally, making the two API’s incompatible. As a bare minimum, support for voice recognition via the Kinect microphone array should be supported as well as speech synthesis. Research into alternatives to SAPI mainly looked into open source alternatives, due to the unique nature of the product. Viable solutions are the eSpeak [7] and Voce [8] projects. Voce has is a Java based project; however it has C++ bindings whereas eSpeak is a C++ library.

A study where using voice based commands for performing simple tasks in a word processing application saw a 12%-30% improvement of performance [9], as the user did not need to move their hands away from the keyboard to the mouse to perform the same task. The same principle applies within this project, as some tasks (such as changing the current editing tool) can be performed in parallel to moving about the three dimensional environment via hand gestures. Ben Shneiderman [10] explains the limitations of voice based input in comparison to hand based input devices. Pointing out that since speaking commands consumes cognitive resources in a human, it makes it difficult to speak whilst solving problems at the same time. Meaning any voice based commands should be simple and intuitive.

## Design and Implementation

### Programming Languages and SDKs

## References

[1] http://subversion.apache.org/

[2] http://cppcheck.sourceforge.net/

[3] http://www.stanford.edu/class/ee368/Project\_11/Reports/Tang\_Hand\_Gesture\_Recognition.pdf

[4] H. Bay, T. Tuytelaars, L. Van Gool. “SURF: Speeded Up Robust Features”, Lecture Notes in Computer Science, 2008.

[5] http://iss.bu.edu/data/jkonrad/reports/HDTT11-04buece.pdf

[6] http://www.cs.sunysb.edu/~algorith/implement/DPsimp/implement.shtml

[7] http://espeak.sourceforge.net/

[8] http://voce.sourceforge.net/

[9] Karl, L., Pettey, M., and Shneiderman, B. Speech versus mouse commands for word processing applications: An empirical evaluation. Int. J. Man-Mach. Stud. 39, 4 (1993), 667–687.

[10] http://dl.acm.org/ft\_gateway.cfm?id=348990&ftid=5995&dwn=1&CFID=328401585&CFTOKEN=28691511

[11]