1,001,001 Faces

A Configural Face Navigation Interface

by

Tzu-Pei Grace Chen

B.A., The University of British Columbia, 1997

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE STUDIES

(Department of Computer Science)

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

June 30, 2003

Department of Computer Science

The University of British Columbia
    Vancouver, Canada

Date _____ July 7, 2003 _____

Abstract

Conventional face navigation systems focus on finding new faces via facial features. Though intuitive, this method has limitations. Notably, it is geared toward finding distinctive features, and hence, does not work as effectively on "typical" faces. We present an alternative approach to searching and navigating through an overall face configuration space. To do so, we implemented an interface that shows gradients of faces arranged spatially using an n-dimensional norm-based face generation method. Because our interface allows users to observe faces holistically, facial composition information is not lost during searching, an advantage over face component methods.

We compare our gradient based face navigation system with a typical, static, slider-based system in a navigation task. Then we compare it with a hybrid dynamic slider system. Results from our first pilot study show that our method is more effective at allowing users to concentrate on face navigation when compared with a static slider interface. This is helpful for face matching tasks as it reduces the number of times users must re-examine faces. Results from our second pilot study suggest that our interface is slightly more effective in coping with correlated navigation axes when compared with a dynamic slider interface. Our third pilot and the formal experiment confirm that while slider-based interfaces are more suited for converging to proximity to the target face, gradient-based

interfaces are better for refinement.

While it may be counter-intuitive that sliders, which are commonly used as interfaces for colour navigation, are inadequate for face matching tasks, our results suggest that new interfaces, such as our gradient-based system and dynamic sliders, are useful for navigation in higher dimensional face space.

Contents

List of Tables

List of Figures

r

Acknowledgements

I would never have considered graduate school if it were not for the inspiration and enthusiasm of my co-supervisor, Maria Klawe. Her dedication to computer science, support for research in Human Computer Interaction, commitment to increase the number of women students in information technologies and skill as an administrator are awe inspiring. She gave me the courage to break out of my shell, the freedom to find a thesis topic that suited me and the hope that with enough enthusiasm and persistence, anything is possible. Of special importance to me was her understanding of the dilemmas and culture shock a woman computer science student, having a non-traditional computer science background, goes through. She provided just the right support and guidance that I needed to get through the tough times of self-doubt and uncertainty. For the lessons of handling my own failure as part of my personal growth, I am most grateful.

I thank my co-supervisor, Sidney Fels, for the opportunity to thrive in his nurturing Human Communications Technologies Laboratory. His unique creativity, unconventional perspective and wholehearted dedication to research made every problem interesting. I am always amazed at how he can make a complex problem seem so easy and solvable. Although he pushed hard sometimes, he always left me with lots of encouragement and motivation to go on.

I appreciate his good temper during our wrestling of ideas, his skill in steering me away from pitfalls, and his diverse background in arts and sciences. His multi-dimensional influence is present in many parts of this thesis.

I would like to acknowledge the generosity and assistance of Singular Inversions, especially Andrew Beatty and John Leung. Without their support, enthusiasm and willingness to allow me to use the FaceGen Software Development Kit, this research would never have been possible.

I would also like to thank Brian Fisher for many insights and references on face recognition and psychology experiments. His suggestions were invaluable. Having him as the second reader of my thesis strengthens the credibility of this face navigation research.

Finally, I would like to thank my parents for making me who I am.

Chapter 1

Introduction

In Kyoto, Japan, there is a temple called Sanjusangen-do that houses 1,000 life-sized Buddhist statues and one very large Buddha figure in the centre. Each statue has a unique face. The belief among the Japanese is that it is possible to find all the faces of their lost relatives among the sculptures. This raises a number of interesting questions: first, given a finite number of statue faces, how is it possible for most Japanese (in a country of 150 million) to find their relatives there? Is it possible that there are only 1,000 Japanese facial "types"? If the faces viewed include all faces in the world, is there perhaps a threshold beyond which we cannot distinguish any more faces? Moreover, what is an effective interface for searching and finding faces?

1.1 Colour-wheel-like Interface for Configural Face Navigation

Our research focuses on the last question. In particular, we are interested in navigating faces by looking at whole faces (similar to the way the temple visitors would), rather than by piecing together facial features. We believe this approach provides a different perspective for face retrieval and could complement existing

methods.

With this motive, we designed a dynamic "colour-wheel-like" interface which allows users to focus on a spectrum of morphed faces at every face space location they select. We call this a "wheel" interface (see Figure 1.1). (The faces on the wheel are generated by a program, *FaceGen*, developed by Singular Inversions [6]. Chapter 4 provides a detailed description).

For the purpose of comparison, we further create two slider-like interfaces (static and dynamic) that allow users to search the same space with a familiar slider mechanism. The wheel and two slider interfaces are tested and compared for navigation effectiveness in our experiments.

1.2 Overview of Experiments and Findings

We conducted three pilot tests and one formal experiment. Each pilot test contributes to our design of the final experiment. Their purposes and outcomes are summarized as follows.

1.2.1 Pilot 1

We ran the wheel interface against the static slider interface varying the number of navigation axes to either three or six. The navigation axes are positively correlated in shape and texture principal components which provided a large variety of faces. We devised a face space of 15,625 precomputed morphed faces in which subjects navigated. Our results show that with merely three navigation axes, there is very little variation in subjects' navigation manners. We also
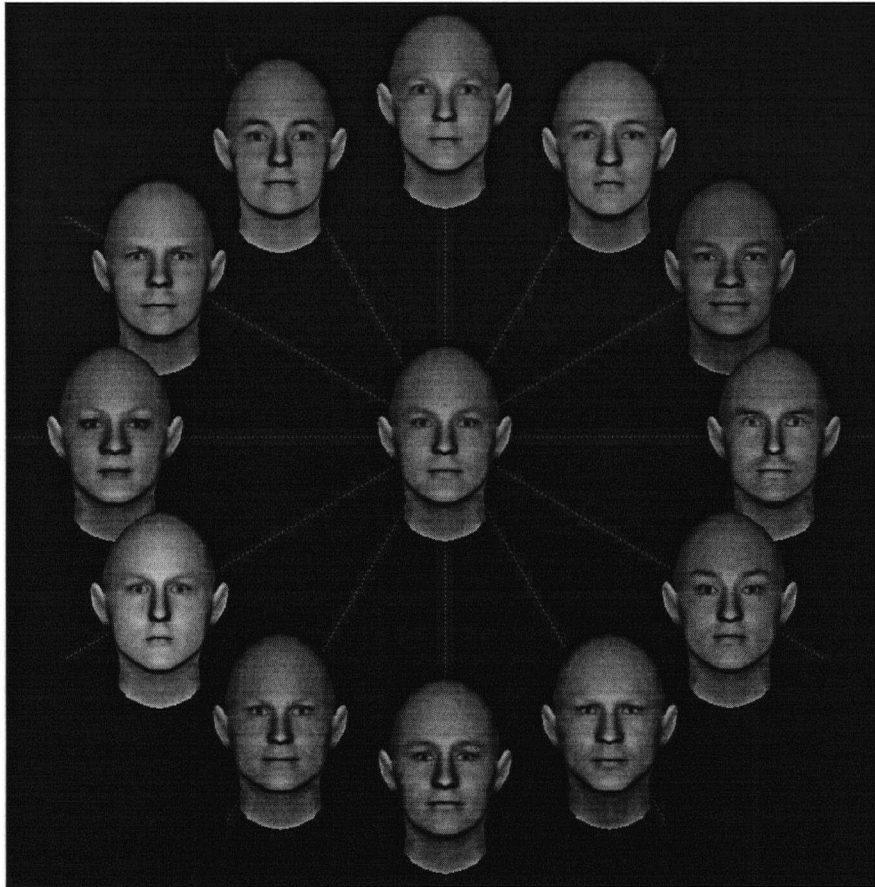
Figure 1.1: The Wheel Face Navigation Interface used in Pilot 1 experiment. Users select the face that appears closest to the face that they are looking for. When a face is clicked on, it becomes the centre face and the wheel is redrawn with a new perimeter of faces, which are one step away in each direction.

found that subjects exhibit a large amount of trial and error behaviour while using the static slider interface; hence, static sliders were replaced by dynamic sliders in subsequent tests.

1.2.2 Pilot 2

Since subjects have shown little variation when navigating with three axes, we fixed the number of navigation axes at six, and compared the wheel interface against the dynamic slider interface. There are two sets of navigation axes. The first set includes three texture and three shape axes that are independent of one another. The second set includes three pairs of correlated texture and shape axes (of which a positive correlated axis and a negatively correlated axis make up a pair). The two kinds of axes call for two separate face spaces to be generated, totalling 31,250 precomputed faces. Outcomes of this pilot test indicate that subjects navigated slightly better with independent navigation axes as they are likely confused by the correlated axes. Thus, independent navigation axes are used for the third pilot test and the final experiment.

1.2.3 Pilot 3

In this experiment, we explored the effect of resolution and the target distance by fixing the other variables. We fixed the number of axes at six, the type of axes set as uncorrelated, and the two interfaces for comparison were the wheel interface and the dynamic sliders. Resolution is either coarse or fine, and the distance of the face target is either near or far from the origin. Faces are

generated during run time, and hence, no precomputation is necessary. However, a total of 531,441 faces would be the population of the entire face space used in this experiment.

This pilot test reveals two interesting trends, which we further investigate in the final experiment. The first trend is that subjects have the most difficulty in finding near face targets under the fine resolution condition. This is counterintuitive as we expect far face targets to be more difficult to reach. The second trend is that subjects visit more places in face space, which are within a two-step neighbourhood of the target faces, when using the dynamic slider interface than using the wheel interface. This trend suggests that although the slider interface has the advantage of allowing users to move more than one step at a time, it is not as useful for refining a face match as the wheel interface.

1.2.4 Final Experiment

Noticing the two trends from Pilot 3, we ran the same experiment with more subjects. Our results do not show the first trend mentioned in Pilot 3, but the second trend persisted. The apparent degradation of the subjects' performance under the fine resolution conditions also suggests that when the distinctiveness of face gradients is below the just noticeable differences of the subjects, the wheel interface is not as helpful. Nevertheless, the data still support the usefulness of the wheel interface for face match refinement purposes, and the slider interface for finding the region of close facial proximity.

These experiments provide useful insights on the benefits and limitations of

the three interfaces we investigated. Results imply that unlike colours, which are usually presented by three variables, face space may not be adequately navigated via a slider interface alone, as faces have a much higher dimensionality than colours. Navigating in face space using sliders may be awkward when the dimensionality is high. We are also likely to be more attuned to faces than colours[1]. This makes the number of navigation controls large, hence the representation of faces and the exploration interfaces in face space are non-trivial research problems.

The structure of this thesis is outlined below.

1.3 Outline of thesis

We present background research, implementation and user testing in this thesis. In Chapter 2, we give an overview of work related to the face retrieval problem by well-known researchers in machine vision, computer graphics and psychology. There is an abundance of scientific research in this area because of people's fascination with faces. Following this, we speculate in Chapter 3 on the usefulness of a colour metaphor for representing faces. A number of interesting relations between colours and faces advocate both for and against the idea of treating faces as colours in research. Although some of the colour and face relations offer a few similarities, we are careful to observe their differences and use these as warnings of potential pitfalls. These relations serve as a key motivation for our

---

[1]If this can be proved to be true, it is not surprising. From an evolution point of view, being able to recognize one's own kind is a vital skill. Furthermore, colour is a subset of faces; therefore there are good reasons why we should be more attuned to faces than colours.

approach, thus there is a strong influence of colour in this thesis.

After speculation of previous work relating colours to faces, Chapter 4, on implementation, outlines the details of our general system architecture. It narrates the design decisions and considerations that made it what it is. It also includes subsequent development on the interfaces used in user testing. Next, Chapter 5 provides an overview of the experiments carried out. It lays out the skeleton of user testing while Chapters 6, 7, 8 and 9 detail the body of the experiments. We follow the convention recommended by [39] and [50] in reporting our experiments.

Finally, important findings are summarized in Chapter 10 where we also suggest future directions for research in face navigation. Additional information relating to user testing and data analysis is included in the appendices.

Chapter 2

Background

2.1 Importance of the Human Face

The human face is a crucial part of the human anatomy. In the context of face recognition, it serves three purposes: identification, non-verbal communication and attractiveness [2, 40].

We often identify people by their faces. As we grow, our faces grow with us, retaining facial features already prominent in childhood. These features are keys for recognition. Research has shown that babies can recognize their parents [9] and imitate their facial expressions [41]. Research has also shown that human brains have a strong ability to process patterns, especially those of human faces [13]. Although faces do not uniquely identify us, they are useful patterns for recognition, as our facial features and colours help indicate our race and origin [40].

The human face is probably the primary channel for conveying emotion through gesture. Our facial expressions can invoke facial expressions in others. For example, a smile provokes a smile. We even communicate through facial gestures. "We 'encode' messages in our own facial expressions, and we simultaneously 'decode' the faces of the people around us" [2]. Listeners likely

interpret speech, for example, differently if the speaker displays joyful or angry facial expressions while speaking. Our eyes may reveal shielded emotion, even thoughts. Our mouths provide the visual complement to our verbal utterance, and our expressions project our emotions [2]. Faces thus play an important role in conversation, and the importance can even be demonstrated by people's frequent use of facial expressions, such as :-) and :-(, in e-mail messages [14].

Finally, faces are probably the most visible "art objects" we present to project ourselves. People of many cultures enjoy personalizing their looks by using cosmetics, facial hair, facial piercings and tattoos to make statements about themselves. Facial decorations thus help to enhance attractiveness in a culture and to mask undesirable features.

2.2 The Face Searching Problem

A common way to find a specific person within a population is through the identification of his face from photographs or individuals. If such information is not available, face composites or sketches can be made based on witnesses' accounts. This procedure, computerized or not, often includes the process of identifying the distinctive facial features of the person. As a result, face composites are produced. Such a procedure is a *component* retrieval type as it involves the piecing together of a face from various facial parts. Figure 2.1 is a crude example.

Figure 2.1: The face composite above is a very crude example, resembling a face collage. Facial features are pasted together rather than blended, making the face rather unrealistic.

2.2.1 Component Methodology

There are several drawbacks to component retrieval methods. First of all, it relies on finding distinctive features. This works effectively for distinctive faces, but less so with typical faces—that is, faces without any prominent features. Second, component navigation encourages the verbal categorization of facial features because face distinctiveness is more conveniently encoded in components[60]. In the context of eyewitness testimony, verbal categorization is one potential factor that can lead to misidentification. Witnesses' descriptions of suspects are influenced by their opinions and emotional state, which, like memory, are changeable. Out of zeal to help solve a crime, one may unintentionally exaggerate a description of a suspect. This, in turn, warps one's impressions of a suspect, and can affect one's proper identification [19]. Loftus, who is a psychologist, states that, "People want to see crimes solved and justice done, and this desire may motivate them to volunteer more than is warranted by

their meager memory. The line between valid retrieval and unconscious fabrication is easily crossed"[35]. Finally, segmented facial features must be united in some way to form a convincing face. Conventional face composite technologies that use component methodology, such as *FACES*[1], *Identi-Kit* [31], *SpotIt!* [10], *FACETTE*[2] and *Imagine*[3], do not produce high quality face reconstruction, despite the use of blending and morphing to unify facial parts. Consequently, the crude results tend to look "criminal". Moreover, these programs have difficulties classifying facial features in their libraries. For example, how best to describe an eye? An eye can be described by its colour, curve, length, width, number of eyelashes, roundness of iris, and so forth. The list goes on, and the possible categories are endless. In fact, many psychologists agree [61] that the number of facial feature categories is so large that it might as well be considered infinite; thus, defining a useful facial classification system is not a simple task.

As an example, Figure 2.2 provides a screen shot of the *SpotIt!* system [10]. Although the face composite is less "criminal-looking" than Figure 2.1, two-dimensional morphing is inadequate to provide a detailed seamless face composite. *SpotIt!* avoids the verbal categorization of facial features, but has difficulty in presenting a concise set of significant facial features to users. Figure 2.2 shows that, for the eye feature alone, seven sliders are provided for manipulation. If one feature requires seven sliders, the total number of controls needed

---

[1]The *FACES* company website no longer exists but it can be purchased via the company called *EVIDENT* at *http://www.evidentcrimescene.com/cata/sketch/sketch.html. (Last accessed April 10, 2003)*

[2]*FACETTE* is at http://www.facette.de. (Last accessed April 10, 2003)

[3]*Imagine* can be purchased at http://www.mshel.com/ima.html. (Last accessed April 10, 2003)

to cover all kinds of facial features can be overwhelming. Our interface design

sidesteps the issue by taking the *configural* approach.
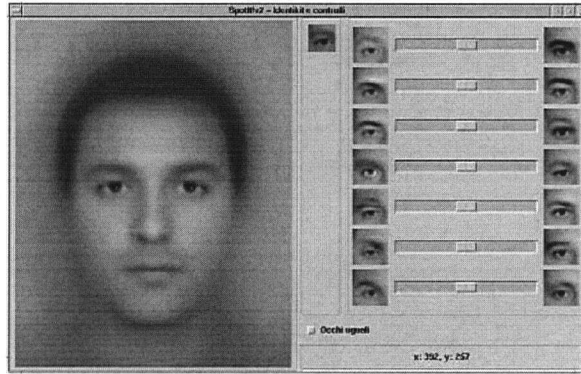
Figure 2.2: This is a snap shot of the *SpotIt!* interface. Although the face
composite is not as crude as Figure 2.1, 2-D morphing can result
in a blurring of details and hence, loss of information. *SpotIt!* is
well-designed in that it avoids the verbal categorization of facial
features, but it remains hard to avoid multiple slider controls of
facial features.

2.2.2   Configural Methodology

Configural face retrieval, contrary to the *component* approach, tackles face

searching by focusing on the arrangement of facial features. It zooms in on

the global layout of each face, and has the potential to compensate in areas

where component retrieval fails. As stated by Hancock et al. [23] and Bieder-

man et al. [7], much psychological research has found the perceptual processing

of face patterns to use holistic analysis rather than by decomposition into local

features. This further supports our design choice for approaching faces config-

urally[4].

---

[4]Note that it may appear that we are equating "configural" with "holistic". For further

In avoiding labelling facial feature categories via the configural approach, it is also worth noting that *SpotIt!* presents sliders of different configurations for the eye. The configural information on the eye and its surrounding features such as the eyebrow suggest that a hybrid of component and configural methods can be useful. This supports our concept that component and configural approaches to face retrieval complement each other. Seeing that *SpotIt* adopts a global component approach with local configural fine-tuning, it is conceivable that its opposite approach—global configural approach with local component fine-tuning—may also be a potentially effective methodology for face search. Combining both approaches and harvesting benefits of both may provide an effective search method.

2.3 Face Retrieval versus Face Synthesis

Most research that involves computing faces can be grouped into two categories: face retrieval and face synthesis. Face retrieval (including face recognition, gesture recognition and face detection) is important in Robotic Vision research area, where realistic 2-D images of faces are used. It also has applications in police work and other areas where automatic detection or recognition of people is needed. Face synthesis (including face animation, face reconstruction and face morphing), on the other hand, is an active research area in Computer Graphics

---

clarification, we use "configural" to mean the arrangement of facial features. Since the way to perceive facial layout is to perceive faces holistically, it may seem like the two words are referring to the same thing. Some psychology literature, however, use "configural" to refer to "parts-based" treatment of faces which is the opposite of holistic. For lack of a better word, we use "configural" to mean the opposite of components. That is, component refers to analyzing each facial feature while configural refers to assessing the overall look.

because of the need for synthetic humanoid faces in entertainment and simulation. Because these applications require 3-D animations, the faces have tended to be three-dimensional but simplified for fast rendering from any point of view.

As technology advances, the gap between face retrieval and face synthesis narrows. There is a desire to construct faces that appear as realistic as possible so that virtual characters are convincing. There are also advantages to retrieving three-dimensional faces rather than two-dimensional face images, for it is useful to be able to detect faces from an arbitrary view point. Therefore, even if face retrieval and face synthesis aim to address different needs, the techniques developed for one can be applicable to the other. In fact, it is reasonable to speculate that face retrieval may become a quick way to create an arbitrary face, and face creation may be used to devise a face database for retrieval.

More recent developments related to face retrieval and creation have adopted a statistical approach. The use of principal component analysis with facial images results in eigenface-based approaches [59] which, in turn, lends itself to content retrieval systems, such as *Photobook* [48], *SpotIt!* [10] and *CAFIIR* [66] [5]. Principal component analysis has also been applied to face reconstruction [8, 57] and evolution, such as in *EvoFIT* [24], and face dimension quantification [23, 47]. A statistically plausible genetic approach has also been popular. DeCarlo et al. [17] and DiPaola [18] have built systems for creating non-photo-realistic humanoid faces. DeCarlo uses anthropometric statistics to generate faces through the use of variational modelling. DiPaola, on the other hand, uses a genetic algorithm to allow users to select "genotypes" that determine

choices for a number of configural and component facial categories. This approach is similar to *EvoFIT*. DiPaola's application is designed for the game, *The Sims*, and hence, has a simple interface that allows users to easily browse through randomly generated faces and modify their looks through a number of parameter sliders.

Although the use of statistics speeds up face retrieval and models with anthropometric plausibility (from normal distribution), it is not always intuitive as to what kind of faces are found. Thus, face retrieval appears somewhat random to users as there is a lack of direction in face browsing. In other words, there is a lack of meaningful organization of the face space.

Our research, unlike others', has both face synthesis and face retrieval parts. We create realistic-looking morphed faces to populate a face space within which face retrieval takes place, (discussed in Chapter 4). This gives us more control but does not imply that the best approach to retrieve faces is the best way to create faces, and vice versa—even though face synthesis and face retrieval are interrelated.

2.4 Face Navigation versus Automated Face Retrieval

Although we are interested in the configural face approach to the face search problem, our primary focus is on the navigational aspect of face space. There is a difference between computer-automated retrieval and computer-aided naviga-

tion for the purpose of face searching. Both involve finding faces that match a set of requirements, but the former places more emphasis on the faces retrieved, and the latter on the process of retrieval.

Retrieval often involves users providing input and the computer doing all the "sifting" work of face retrieving. This does not provide users with a good sense of the face space. Users may know how large the database is, but not necessarily where faces are located and where to find them again if needed. It is like being blind-folded while transported, users have to place their trust in a machine to retrieve faces close to their specifications.

The second process, on the other hand, provides users more control. Users are in charge in navigation. This naturally involves more cognitive load on the users, but they have more freedom and a greater sense of direction in face space. They are better able to remember the path they took to reach a certain face, and hence have more intuition about what is going on.

Although one could argue that the second process may take too long in a massive face database, and therefore, is not an efficient application for face retrieval, it is important to always consider the human as the centre of the application design. Instead of making the computer mimic human face recognition, it may be a good idea to make use of natural, human face recognition ability—which is the most perceptually effective system we have at this time. This is an important aspect of our research.

In this chapter, we have provided an aerial view of face retrieval research. We point out the importance of faces. We compare component and configural

methodologies in face search problem. We then briefly discuss the difference between face retrieval and face synthesis, and compare and contrast face navigation and automated face retrieval. We argue for a configural approach in face navigation where users face recognition skills are utilized. In the next chapter, we explore a number of relations between colours and faces. Their similarities and differences contributed to the design of our navigation system.

Chapter 3

Colour Space versus Face Space

Colour is a subcategory of face dimensions but face colours (predominantly the skin tones) fall in a narrow colour spectrum. The multi-dimensional nature of both colour and face space is intriguing and is the motivation behind our research.

In face recognition research, one of the largest challenges is to contrive a face similarity metric to represent differences of all the faces in the world. Psychologist have sought such a metric to explain various face recognition phenomenon. but so far, have not succeeded. It appears to be a very difficult problem.

Faces are transient and their dimensionality is considered to be infinite [61]. Our faces change due to natural causes such as age, mood and health, and due to external factors such as cosmetics, surgery and injury. However, face transience does not seem to affect humans' acute face recognition skills. We are capable of recognizing people from a distance; we can recognize people from an angle; we can detect friends in disguise; and we can even call out the names of childhood friends seen for the first time in decades. The multiple dimensions of faces and our acute face recognition ability make designing a face similarity metric a difficult task.

In designing our interface, the lack of a face similarity metric is problematic because we need to create a face space in which to navigate. Without the metric, direction and position in the face space cannot be meaningful. We have, however, found an alternative. Figure 3.1 shows a spectrum of faces created from *FaceGen* [6], which uses a face metric based on principal component analysis methods. Although this is not necessarily the best face similarity metric, it represents the "state of the art" at the moment. Regardless of its merit, we make our interface independent of *FaceGen* and its metric, so that our designs can be applied to other metrics in the future.

3.1 Previous Face Metrics

The eigenface, by Turk and Pentland [59], is a well known face metric derived from the application of principal component analysis to face images. It is essentially a related set of facial characteristics that can be used efficiently by a computer to recognize a person's face. Although the eigenface works well for computational face recognition, this does not imply that it makes a good face similarity metric. In fact, research [5, 46] shows little correlation between eigenfaces and human perception of face similarity. The "shape free face" approach normalizes faces by morphing them onto a standard face grid before the application of the eigenface [23]. Although Bruce et al. do show some correlation between the eigenface and human perception of face similarity, their findings do not provide enough evidence to convince other researchers of its superiority.
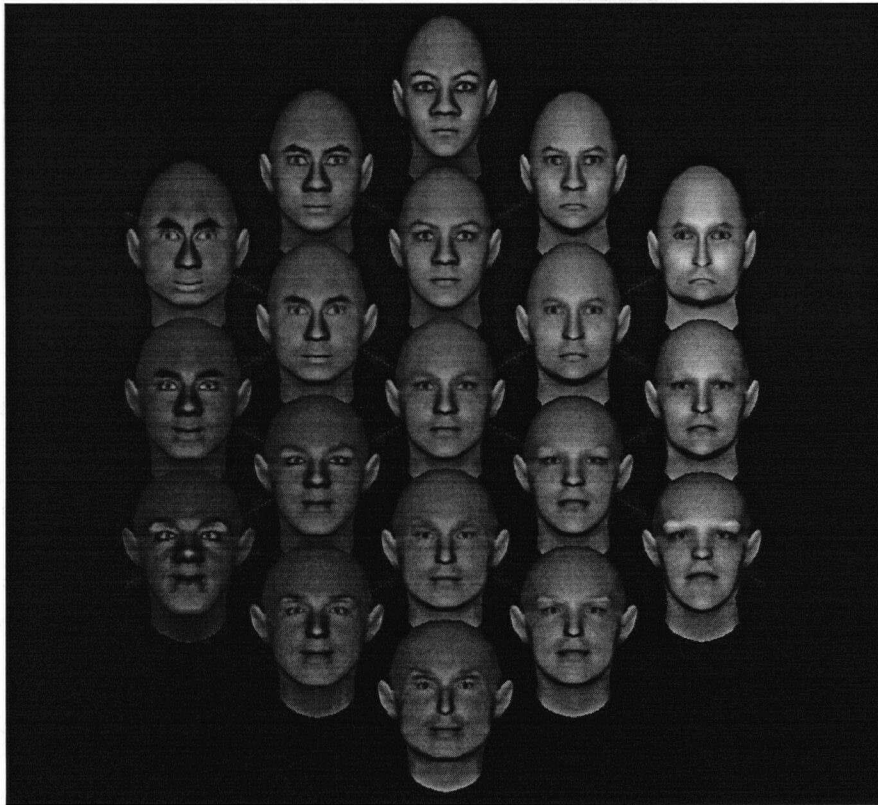
Figure 3.1: The face spectrum shown here parallels a colour wheel.

Another approach was taken by Marquardt, a plastic surgeon who created a face metric for measuring attractiveness [38, 45]. His design is based on the Greek golden ratio (1 : 1.618). He claims that the better the face fits his beauty mask, the more attractive it is. Although Marquardt displays many examples of beautiful faces fitting snugly inside his metric, it is difficult to prove that a ratio determines attractiveness. Many factors make a person beautiful that may not be captured by a function of proportion. Alley et al. [1] points out several. For example, youthfulness alone is considered an attractive quality in

women, because for men, it symbolizes reproductive capacity. However, women may prefer facial characteristics in men that reflect physical strength, as they imply health and the ability to provide resources and protection. Familiarity can also increase the attractiveness of faces, and often, attractive faces are atypical. These are qualities that the golden ratio cannot capture; hence, the usefulness of the beauty mask remains limited.

One approach that has not been explored in seeking a face similarity metric is the use of a colour metaphor. This is perhaps because faces and colours are fundamentally different. However, the colour metaphor is a plausible starting point because both colour space and face space are multi-dimensional. In fact, colour is a subset of facial features. Colour is a well-researched topic. There are many systems representing colours, such as CMYK, RGB and HLS [52], [33]. The design of a similarity metric for faces can perhaps be inspired by colour research, since colours have fewer dimensions than faces, and are, therefore, more manageable in size and easier to grasp conceptually.

In the next section, we present six interesting analogies between colours and faces. Some of the analogies shown in research by [19], [51] and [32], suggest some similar phenomenon occur in our perception of colours and faces. This is encouraging evidence that the application of a colour metaphor may offer a fresh perspective on the face metric problem. However, other research (see [64] [3]) indicates differences between colour and face spaces. Finding a face similarity metric is outside the scope of this thesis. We present an overview of the research exploring colour and face relations so that in understanding the

difference between them, we can avoid the pitfalls of over-using the framework already established in colour research. The existing literature not only serves as motivation for our research, but also allows us to speculate as to how far the colour metaphor can carry us in understanding faces.

3.2 Colour and Face Relations

There are a number of interesting relations between colours and faces from the point of view of psychology, neuroscience, law, genealogy and anthropology.

3.2.1 Colour in Face Recognition

Colours and faces are both perceived holistically. When we look at colour print-outs, we do not detect the small, primary coloured pixels. Nineteenth century Impressionist artists such as Renoir, Van Gogh and Monet realized that human perception blends the dots and strokes of primary colours to form secondary colours. This phenomenon dominated their style of painting. Similarly, we perceive faces holistically. We do not blend the facial features together like colours, but we read facial features together to see the whole face. In fact, in comparing face recognition and object recognition, researchers notice that the context of the face facilitates detection of the differences of individual facial features (such as nose and mouth) [7].

Yip et al. [67] notice that contrary to past research, which suggests colour confers little recognition advantage for identifying people, colour actually contributes significantly to face recognition, especially when shape cues are de-

graded. Skin colour histograms are also used in face detection and tracking by Kawato et al. [28]. These findings suggest that it is worthwhile to include colours to solve face retrieval problems. Although many commercial face retrieval systems only provide black and white images, our system provides colourful, realistic-looking, morphed three-dimensional faces.

3.2.2 Colour-blindness and Face-blindness

Colour-blindness is defined as the inability to distinguish between some colours with the same intensity. It is an inherited condition, sex-linked and recessive. Consequently, very few women are colour-blind, but approximately one in ten men has some degree of colour-blindness. Symptoms usually involve patients having difficulty distinguishing shades of a particular colour, but in most cases, symptoms are mild.

Similarly, one can be face-blind, a condition called *prosopagnosia*. Prosopagnosia, or face-blindness, is a neurological condition that renders a person incapable of recognizing faces (including her own) [42]. It arises when the part of the brain dedicated to face recognition becomes damaged, but this is unrelated to a person's ability to see faces. Patients cope by relying on other visual clues, such as the person's walk, clothing and voice so that their object recognition skills can assist in face recognition. There are other face associated disabilities

such as Capgras syndrome[1], Frégoli syndrome[2], Asperger syndrome[3] and Mobius syndrome[4] but they are less relevant than colour related conditions. (For additional information on face-related illnesses, refer to [16] and [68].)

Although colour-blindness is an optical disability and face-blindness is a neurological damage, our brain may be processing colour and faces in similar manners. In raising awareness of these disabilities, we anticipate our system, which takes on the configural approach to navigating colourful morphed faces, to have less success with prosopagnosic and colour-blind patients. However, there are remedies. Providing colour-blind users with black and white images may avoid the problem. Offering component face retrieval mechanisms can also help prosopagnosics to use their object recognition skills in identifying faces. This further supports the combined use of component and configural methodology.

[1] Capgras syndrome is a neurological syndrome often mistaken for insanity. Capgras' patients typically identify people close to them as being impostors—identical replicas of the people they know. This is because the patients can recognize faces but cannot find the emotions associated with them. Classically, the patient accepts living with these impostors but secretly "knows" that they are not the people they claim to be [68].

[2] Frégoli syndrome is the delusion that one person has taken on a different appearance or several identities. This syndrome is named after an Italian actor known for his skill in impersonating others. The defining symptom is that the affected patients misidentify strangers as familiar (often famous) people whom they believe to be persecuting them. They do not mistake the appearance of the strangers, but think the strangers' apparent non-resemblance to be some deliberate disguise [68].

[3] Asperger syndrome is a mild form of autism which affects how a person communicates and relates with others. Patients often have limited gestures as well as poor comprehension of non-verbal communications, such as facial expressions. However, many exhibit exceptional skill or talent in a specific area and are often obsessed with complex topics such as patterns, music and technology [30].

[4] Mobius syndrome is a rare disorder characterized by lifetime facial paralysis. People with Mobius syndrome cannot smile or frown, and they often cannot blink or move their eyes from side to side. These symptoms can severely limit patients' social interaction with other people as non-verbal communication is limited [16].

3.2.3 Colour and Face Opponent Mechanism

As Hurlbert indicates in [27], our senses are "optimally designed to signal relative changes against the status quo, not absolute value". Just as there is evidence of colour-opponent mechanisms, there is evidence of face-opponent mechanisms. An example of a colour-opponent mechanism is the phenomenon that occurs when one "sees" a patch of green on an all white surface, after staring at a patch of red for some time. This is due to nerve fatigue in the eyes, which causes a "displacement" effect—hence white is perceived to be green. The face-opponent mechanism is shown in the experiments by Leopold et al. [32]. They discover that exposing a subject to an individual face for a few seconds can significantly affect his subsequent perception of face identification. In fact, subjects who have been exposed to the "anti-face"[5] of a face stimuli have the tendency to mistake the average face to be that stimuli (see Figure 3.2). This is intriguing as the average face lies precisely in the middle of the face stimuli and its anti-face, suggesting the after effect of an opponent mechanism.

This finding seems to support the idea that our brains may process colours and faces in similar ways. However, we need to be aware that the creation of an "anti-face" is purely due to principal component analysis. It is a statistical way to provide a structure to a face space. It is most likely that we do not think in terms of positive and negative faces when we recognize people. Similarly, there are no "anti-colours." There are complementary colours but a colour's

---

[5]An "anti-face" is essentially a face which is the opposite of the face stimuli. If the face stimuli is fat, then the anti-face is thin; if the face stimuli has a narrow nose, then its anti-face has a wide nose. This is achieved by Blanz et al. [8] who applies principal component analysis to hundreds of aligned 3-D face scans to produce seamless, three-dimensional, face morphing.
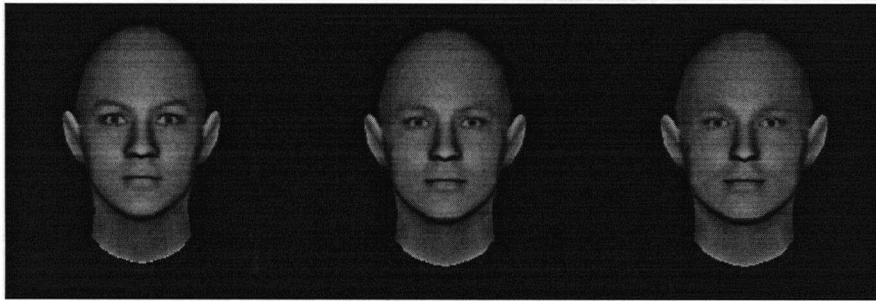
Figure 3.2: A face opponent mechanism is shown by Leopole et al. by training subjects to be familiar with a face, such as "Mr. Red" on the left. After this exposure, subjects quite often mistake "Mr. Neutral" in the middle to be "Mr. Green" on the right. "Mr. Red" and "Mr. Green" are anti-faces of each other; that is, their features tend to look opposite. In this case, "Mr. Red" has a wide nose and thick eyebrows but "Mr. Green" has a narrow nose and light eyebrows. "Mr. Neutral" has features that balance the two.

complement is not its negation. Despite this difference, the fact that people are better at sensing relative changes, as Hurlbert [27] points out, suggests our incorporation of face gradients in our interface to be a useful feature. Anti-faces also provide useful visual cues to direct users away in the opposite direction if desired.

3.2.4 Verbal Effects on Colours and Faces

Schooler et al. [51] and Dodson et al. [19] have shown that memory performance for face recognition is impaired if face stimuli are described verbally after subjects are exposed to them. Researchers attribute this result to the shift from the visual processing of faces to the verbal decoding of faces. Because we are better at describing faces by component, verbal descriptions of faces often exclude facial configuration information, which is hard to articulate. Consequently, our

tendency to rely on a verbally-biased recoding of what we see interferes with our original visual memory of the face.

Similarly, Schooler et al. [51] show that the same effect can occur with colour recognition, for similar reasons. Colour descriptions are hard to put into words, as most people are only familiar with the names of primary and secondary colours. Thus, verbal recoding of the original memory of colour can interfere with subjects' ability to apply their original visual impression.

In light of this, verbal categorization of faces is avoided in our design.

3.2.5   Colours, Emotions and Facial Expressions

We often associate feelings with colours. Blue gives us a sense of sadness, and red gives us a sense of anger [21]. Our emotion is best conveyed by our facial expressions, and Plutchik [49] cleverly uses a colour-wheel model to form a multi-dimensional model of emotions.

In this model, there are eight primary emotions: anger, fear, acceptance, disgust, joy, grief, surprise and expectation. Within them, anger is the opposite of fear, joy is the opposite of grief, acceptance is the opposite of disgust and surprise is the opposite of expectation. They are arranged by their intensities, just like the hue saturation of colours. When you mix primary emotions, you get mixed emotions. For example, $joy + anger = pride$ and $fear + surprise = alarm$. Mixing more emotions becomes harder to describe in words, just as it is more difficult to describe colours of varying degrees of hue, tone and saturation. Although emotion is not directly related to our research because we are only

concerned with navigation in perceptual face space, facial expression is one of the factors in face recognition. If emotions can be depicted by facial gestures, then it is conceivable to replace Plutchik's emotion model with facial expressions, forming a subset of facial dimensions. The colour navigation interface may therefore be applied to this subspace.

In contrast, Ekman codes facial expression by using the facial muscles in his Facial Action Coding System (FACS) [20]. Because we have a finite number of facial muscles (44 of them), it is possible to codify emotion in some physical way. Facial appearance (with a neutral expression), however, is more transient. One's appearance can be changed by many factors (as described in the introduction of this chapter). Facial muscles underneath the skin alone may not yield a good method of quantifying faces because it is the facial surface we are concerned about. Nonetheless, muscles and bone structures may offer ways to code valuable configural information.

3.2.6 Primary Colours and "Primary Faces"

Colours have a primary category which includes red, green and blue in an RGB colour system. Do faces have a primary category of faces? The answer is "yes and no" depending on your point of view.

Recent discoveries in human DNA suggest that all humans can be traced back to common ancestors [3], [64]. In other words, a biological "Eve" and "Adam" may have existed long ago. Unlike the biblical "Adam" and "Eve," these individuals may not be the only people living at the time but are thought

to be the lucky ones who survived to pass on their genes. These conclusions are triggered by the finding that humans have less genetic variation as a species than our closest animal relatives, the chimpanzees. To trace our lineages, scientists investigate mitochondria DNA, which gives vital clues [3].

Mitochondria is a tiny cellular structure, the "power house" of every cell in the human body. When a human egg and a sperm come together during conception, the mitochondria of the cells do not merge, unlike the chromosomes that form a new human being. The mitochondria of the sperm cell wither and die, while the mitochondria of the egg cell live on. In this way, all children inherit their mitochondria DNA from their mothers, and it is passed on intact through the female line. Because mitochondria DNA mutates[6] over eons, scientists are able to identify branches of human ancestors. If two different people share the same mutation in their DNA strands, they must share the same maternal ancestor in whom this mutation took place. Assuming that genetic mutations occur more or less regularly over time, scientists can compare two samples of mtDNA, noting where they share mutations and where they do not, and estimate the time when the individuals' ancestral populations diverged. Similarly, analyzing the y-chromosomes gives scientists a way to follow the paternal line of the humans [64], as y-chromosomes pass from father to son without genetic re-shuffling.

This discovery, on one hand, may allow us to argue that yes, primary faces do exist, namely those of the biological "Adam" and "Eve" from whom humans

---

[6]Mutation is a random change in a DNA sequence which is the result of "copying mistakes" during the process of cell division. It occurs at a rate of around thirty per genome per generation [64].

are thought to have descended from. However, it is hard to believe that only two faces could be responsible for all the diversity in the world. One would expect a set of primary faces to include representatives of races such as *afer* (African), *americanus* (Native American), *asiaticus* (East Asian), and *europaeus* (European) as classified by the eighteenth century Swedish botanist, Carl von Linne [64]. Even this might not be sufficient since there are other factors such as our environments and natural selection which make us look different and unique.

Our looks are determined by both nature and nurture, so knowing the genetic make up of a person does not necessarily give the full picture of that person's facial appearance. In fact, Oppenheimer [3] states that we all look different because we live in different environments, when we are really the same under the skin. People who are from hot-climate areas have darker skin than people who are from cold-climate areas because greater amounts of skin-darkening melanin is produced to protect the body from the sun [54]. People who have more calcium in their diet grow taller. People who eat a lot and exercise little have rounder faces due to their body fat content. Even identical twins do not look exactly alike. So, why does each race have certain "typical" looks? It is believed, as geneticist Spencer Wells explains, that this is due to "localization of mating habits" [64].

Before travel became so common, people living in the same region looked similar to each other over time because of similar living conditions, as well as the tendency to choose mates from those living close by. In addition, it was also

common in "the old days" for cousins to marry; therefore, some genes are shared. Over time, this "inbreeding" produced a distinctive look within regions. Today, in addition to the effect of frequent travel and migration, inter-racial marriages are increasing. It is, however, not yet clear whether these trends will have a homogenizing effect or diversifying effect on the human face.

In view of this, we might also argue that there cannot be primary faces because there are too many factors involved in our development. This may be another reason why finding a face similarity metric is such a hard problem. Perhaps the only reasonable conclusion to draw from this might be that primary faces could exist only if there were not affected by environmental factors or mutations.

Having discussed this issue from a genetic point of view, we need to clarify that the concept of "race" is actually not based in biology. We are one species with different "racial" characteristics, not separate subspecies. It is possible that there could be primary human faces that run across all "racial groups," only overlaid by specific "racial characteristics". This could explain the "other-race effect" where a racial group one has never seen before looks alike until one gets used to it. Then suddenly, everyone looks like someone one has met before from the racial groups one is familiar with. This is the principle behind Sanjusangen-do. Visitors find the statues all very similar at first. However, closer inspection reveals that the sculptures are all unique. Our interface also provides a similar "feel." The application of face morphing provides gradation of facial changes that entice users to inspect similar faces; and to their surprise,

a closer study of faces reveals dissimilarities.

3.3 Colour for Face?

Given these analogies, we wonder if colour selection interfaces can be used effectively with faces. Colours and faces certainly have some similarities in human perception. As mentioned previously, we read both faces and colours holistically. Human beings can be colour blind and face blind. We distinguish similar colours and similar faces better relatively [27]. A face opponent mechanism can occur like a colour opponent mechanism [32]. Verbal foreshadowing effect can take place with both colour and faces [19, 51]. Lastly, facial expressions convey emotions which can be arranged on a colour wheel [49]. However, faces do not seem to have "primary faces" or face prototypes like colours [3, 64]. This makes our desire to imitate colour-picking interface difficult but not entirely impossible. Readers will see how this is done in Chapter 4.

In order to develop an effective face navigation interface, we mimic two colour-picking interfaces. They are: the slider interface for continuous colour spectrum (see Figure 3.3) and the colour swatches for discretized colours (see Figure 3.4). We are curious to find out what happens when we apply these two interfaces to faces. We thus implement and test them both. However, due to computational limitations, we "down-size" the colour swatches idea as we cannot display a large number of realistic-looking faces real-time. Nevertheless, the discretized colour spectrum nature is preserved. Our investigation is unveiled

later on in the chapters on implementation and experiments.
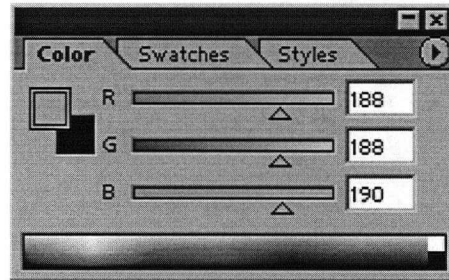
Figure 3.3: The RGB colour sliders in Adobe Photoshop for colour selection. Users can manipulate the sliders for red, green and blue to obtain their desired colours. They can also click randomly on the continuous spectrum at the bottom.
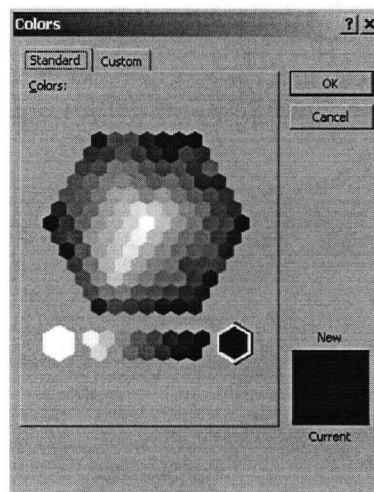
Figure 3.4: The discretized colour hexagon in MicroSoft Word application. Users select individual hexagon patches to customize colours of their words.

In this chapter, we have touched on the challenge of creating a face similarity metric. We highlight a number of relations between colours and faces arguing their validity in support of a "colour" treatment of faces. Whereas the previous chapter provides a literal review of the face search problem from a computing science perspective, this chapter provides analogies of colours and faces from psychology, neuroscience, law, genealogy and anthropology.

Problems can often be solved by observations from different perspectives. New technologies are frequently inspired by borrowing ideas from other disciplines. However, it is important to check the appropriateness of such imitation. This is what we attempt to do in this chapter by offering a closer look at colours and faces. In the next chapter, we introduce our system. We raise a number of design issues and explain how we implement the navigation interface to overcome them.

Chapter 4

Implementation

As discussed in the previous chapter, there are several parallels between colours and faces. However, the two remain fundamentally different; what works well for colours as an interface may not apply directly to faces. Hence, we need to be conservative in our use of the colour metaphor to faces until more research has been done. What then is the best methodology to adopt? How do we overcome some of the challenges with faces?

To answer these questions, we discuss three important design decisions in the first part of this chapter and unveil our interfaces in the second part. Following that, we describe further development made for the purpose of user testing. Modifications made that are specialized for each experiment are, however, left for later chapters.

4.1 Design Decisions

4.1.1 A Configural Approach

As mentioned before, there is research that advocates a configural approach to faces [7, 23]. We include this strong element in our design as we agree with the Gestalt Principal, which is based on the belief that "the whole is greater

than the sum of its parts" [63]. The fact that we can define parts or even sense individual facial features and verbally describe them, does not necessarily mean that our brains work the same way when processing faces. Presenting holistic facial properties to users should be more effective than presenting segmented facial parts [7, 23].

4.1.2 Use of Eigenface

We also believe there are good reasons to use eigenfaces (principal component analysis) as a face similarity metric. Although there is research arguing against its use, because of its perceptual implausibility [5, 23, 46]; in our case, the advantages outweigh its limitations.

In Section 2.2.1, we explain how the number of categories of a single facial feature, such as an eye, can easily be large. An eye can be described by its colour, texture, size, curvature, number of eye lashes and so forth, but as the list goes on, there are bound to be some qualities that are more essential in describing the eye than others; hence trivial categories may not be useful or required. The advantage of using principal component analysis is that it captures the principal variations of faces, hence lower order principal components that represent trivial facial qualities can be ignored. Even if lower order principal components capture distinctive features, such as a mole, the trend is to discard the lower order principal components to economize the computational complexity of the representation of faces. In *FaceGen* [6], which represents faces using 128 principal components, the face space contains a wide variation of faces that

are presented without significant loss of detail in our face space.

The inability to represent faces using a small set of primary faces is another good reason to use eigenfaces. The higher level principal components act as primary faces, capturing the majority of facial variations. Although not every principal component describes a meaningful quality such as age, skin colour or mouth shape, an additional benefit is that it does not categorize facial components, and hence avoids the problems of foreshadowing resulting from verbal categorization [19, 51].

Despite the disagreement between the way our perceptions work and eigenface representation [4, 23], there are good reasons to adopt this approach.

4.1.3 User Face Recognition

It is also important to mention that our design off-loads the task of face recognition onto the user. Since the human face recognition ability still far surpasses that of machine algorithms, it makes sense to engage users this way. Finally, because of evidence that colour contributes to facial recognition [28, 67], our design includes realistically coloured representations of faces.

We have outlined three important design issues. The reasons behind those design decisions result from careful evaluations of face-related research. These are, of course, not the only implementation considerations. We introduce our system next, tying in metaphors and revealing mistakes made behind the scene. This is followed by descriptions of the two challenger interfaces we build. Subsequent modifications made to our system for user testing are presented in the

next chapter.

4.2 General System Architecture

Taking the limitations of component retrieval into account, our configural face navigation interface is designed while taking into consideration four metaphors: a genetic tree, a colour palette, a map and a web browser. Each metaphor has qualities that assist navigation.

We make several variations of our face navigation systems for evaluation. They are based on configural navigation using *FaceGen*, OpenGL [65] and Tcl/Tk [25] on the Windows 2000 OS, and the key features explored are described below. We also devise various finite and holistic face spaces for different conditions of user testing. Conceptually, our system consists of two parts: the interface that users navigate with, and the face space in which they explore. Figure 4.1 depicts how the system looks. One could imagine the navigation interface slides within this face space to reveal a sample of faces as users navigate.

We describe the face space first, then the navigation interface.

4.2.1 The Face Space

Before our interface can be used, we must make decisions about what face space the users navigate. If the word "navigation" implies a method of determining position, course and distance travelled, then we need to give meanings to direction, position and distance in this face space so that users have a greater sense of where to explore. In addition, it is also desirable to provide visual cues
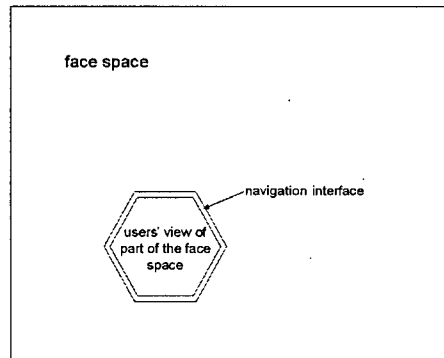
Figure 4.1: We devise a face space and a navigation interface. This is the conceptual view of our system. By using our interface, users view a sample of faces in the face space.

so that users feel as if they are travelling when navigating within our system. Although our navigation interface is tailored for the face space we devised, we are careful to distinguish between the face space and the navigation interface so that the usefulness of our interface is not solely dependent on the face space itself.

We will next describe the face morphing technique used to devise a "norm-based" face space and explain what it means to navigate along the axes in this space.

Face Morphing

Faces and colours are both multi-dimensional. If each unique face has a unique pattern of genes, then mixing the genes results in mixing the faces. By allowing mixing faces to be as simple as mixing colours, we need a way to perform face morphing colourfully and realistically so that we can avoid the pitfalls of

conventional face composite software.

Many psychological studies ([12, 56]), as well as face-composite software, employ 2-D morphing to show combinations of faces. This approach hinges on setting accurate corresponding points of the source images. If not enough accurate corresponding points are available, results can suffer from artifacts that render morphed faces unrealistic and abnormal. Setting corresponding points automatically is itself a difficult problem. Manually setting corresponding points is also laborious. Therefore, convincing morphings are difficult.

An example of the state of the art work on face morphing is that of Blanz and Vetter [8]. Their approach is revolutionary because they produce very realistic-looking morphed faces with linearly combined facial qualities. Their method involves taking 200 face scans of people, aligning the face scans, and applying principal component analysis to obtain a basis that spans a continuous face space. The basis then serves as a set of axes that quantify a unique face as a face vector from the origin (where the average face is positioned). In other words, a distinct face can be produced by warping the average face according to a set of values from the basis.

The Norm-based Face Space

Blanz and Vetter's approach is an interpretation of Valentine's norm-based face space model [60], which he devised to explain the various effects associated with face recognition. Essentially, every face is encoded as a point in a n-dimensional space in this model along dimensions that differentiate faces, such

Figure 4.2: A three-dimensional norm-based face space. The three axes that structure the face space extend out from the central average face.

as gender, age, race or eye colour. The distance between any two points is analogous to the similarity between the two faces. The closer the face is to the average face at the origin, the more typical-looking it is. Similarly, the further away, the more distinctive the face. Although this face space model does not necessarily coincide with our psychological perceptual face space [11, 12, 58], it is conceptually intuitive and has gained strong support in psychology [12].

Figure 4.2 is our simulation of how the space might look in three dimensions.

We consider other models of face space as well, namely the Voronoi model by Lewis et al. [34]. Although this model has a unique way of explaining the caricature effect in face recognition, we feel it does not offer an intuitive face space for navigation.

FaceGen

Similar to the work by Vetter and Blanz, *FaceGen*, a powerful 3-D face morphing program, is created by Singular Inversions. We use the **FaceGen SDK version 2.1** to create faces to populate our face space for navigation. Unlike Vetter and Blanz, the makers of *FaceGen* fit a generic head to each face produced. They also use a total of 128 principal components. These are comprised of 80 shape modes and 48 texture modes (of the shape, 30 are asymmetric, and of the texture, 8 are asymmetric). Figure 4.3 shows examples of faces one can obtain from the four kinds of principal components. Face generation is further optimized by the use of the Intel Image Processing Library and the bypass of zero value face parameters to allow for faster generation during run time for interactive performances.

The Navigation Axes

The 128 principal components of FaceGen serve as the basis vectors for FaceGen face space. Each describes different qualities of faces. These capture the variations of 300 face scans. Since this is a statistical method, we cannot label these
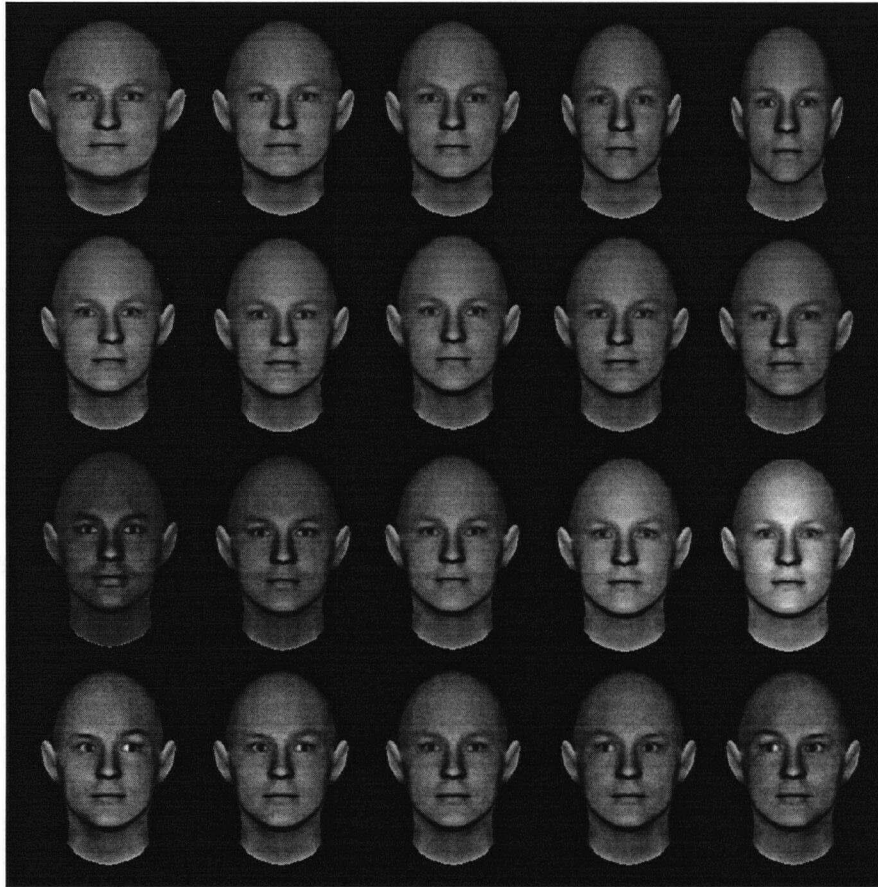
Figure 4.3: *FaceGen* can produce a large variety of faces. The top row shows the faces generated from the first symmetrical shape principal component alone, from a negative value to a positive. The second row shows the faces generated from the first asymmetrical shape principal component. The third row shows the faces generated from the first symmetrical texture principal component, and the bottom row shows the faces generated from the first asymmetrical texture principal component. Notice that the central column is made up of the average face when the value of the principal component is zero. Although we have shown the effect of one of each of the four kinds of principal components, they can be used together to create a combined effect.

as any meaningful facial component or configural properties. We can, however, still use the principal components as navigation axes. These navigation axes can be thought of as dimensions of manipulation as the changing of values along the principal components results in different combinations of facial qualities. We describe our navigation axes in detail in later sections as we apply different principal components (which do not need to be used separately) to create different navigation axes and generate faces for our user studies. Our interface also provides a control panel (Figure 4.5) which allows for the manual selection of desired principal components such as navigation axes. In our attempt to be concise, we also avoid using the word "dimension" to refer to navigation axes, as it may be confusing to readers.

By using *FaceGen*, we have chosen what position, direction and distance mean in our face space. In every position of this space, there is a face whose qualities are defined by its location. Although the *FaceGen* face space is continuous, we discretize ours so that we constrain and simplify navigation for users. Direction in this face space is determined by the qualities captured by the principal component in that direction. Depending on which principal component is used as navigation axes, users can determine what face prototypes to expect by navigating towards that direction. Although the principal components do not provide meaningful terms for the direction of navigation, it is sufficient to show faces to users as landmarks. Distance travelled in this face space is related to the origin of the face space. The further the user is from the origin, the further away they are from the average face, and the less the current face resembles the

average face. Being further away also implies that the faces located there have a higher degree of prominence—even the exaggeration of caricatures.

The seamless morphing also provides users with a sense of movement as they travel in the face space, which is an additional benefit. Although psychologists, such as Busey [12], warn about the use of face morphing in face recognition research (he points out that morphing makes faces look younger by smoothing wrinkles and other details), we believe the 3-D face morphing of *FaceGen* does not limit us. This is because we do not include any original faces that are used to create *FaceGen*. All the faces we generated are "child morphs", and are therefore, similarly "smoothed" in detail.

4.2.2   The Wheel Navigation Interface

Having defined our face space, we next describe our navigation interface. We create a colour-wheel-like interface for our system. A conceptual drawing of the interface design is illustrated in Figure 4.4. We also create two slider-like interfaces for comparing our interface against. The slider based interfaces share a number of modules developed for the wheel interface. Since they are not the focus of our system, we do not explain their implementation in detail.
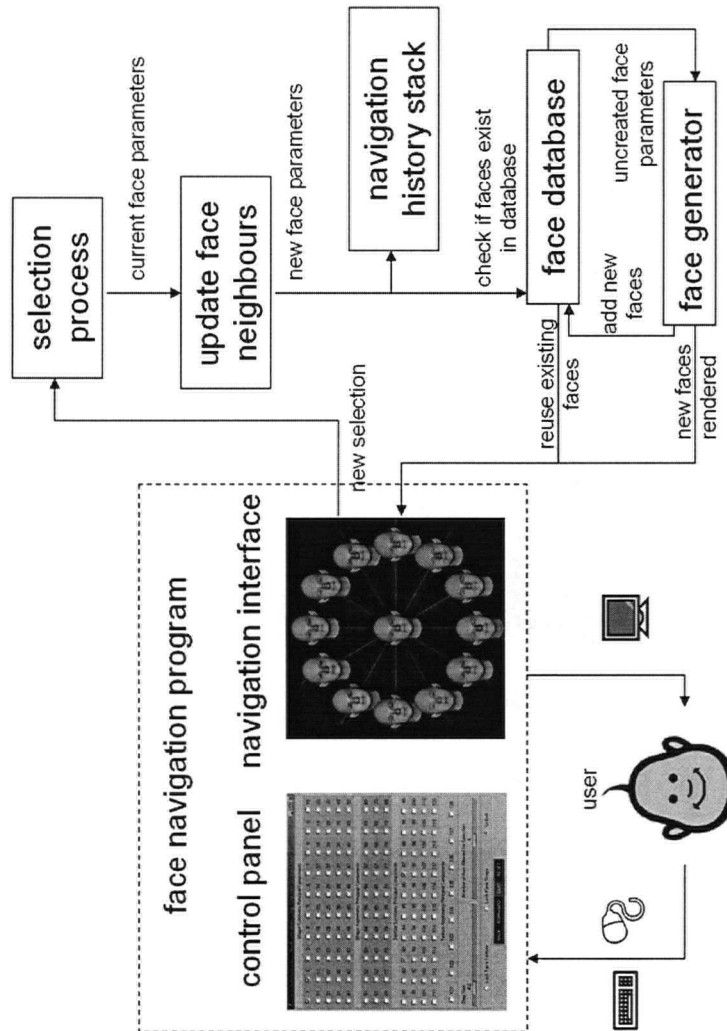
Figure 4.4: Overview of the face navigation interface. Users receive visual outputs from the computer monitor and enter inputs via the mouse and keyboard. They interact with an OpenGL window and a Tcl/Tk control panel. The program handles all the updates in the background, hidden from the users.

The windows seen by users of our system are shown in Figures 4.5 and 4.6. Figure 4.5 shows the control panel (programmed in Tcl/Tk [25]) and Figure 4.6 (programmed in OpenGL [65]) displays the faces selected via the control panel. In this case, the control panel shows a total of four principal components selected (1, 2, 83 and 87). Principal components 1 and 2 are the first two symmetrical shape principal components, while principal components 83 and 87 are the third and seventh symmetrical texture principal components. The step size is set at 4.0 for all axes, and the number of axes to be displayed is set at four.

Because we borrow many ideas from colours, we will focus on how the colour metaphor is applied. We will also describe other mechanisms such as zooming, undo, loading, locking and the control panel in the next subsection.
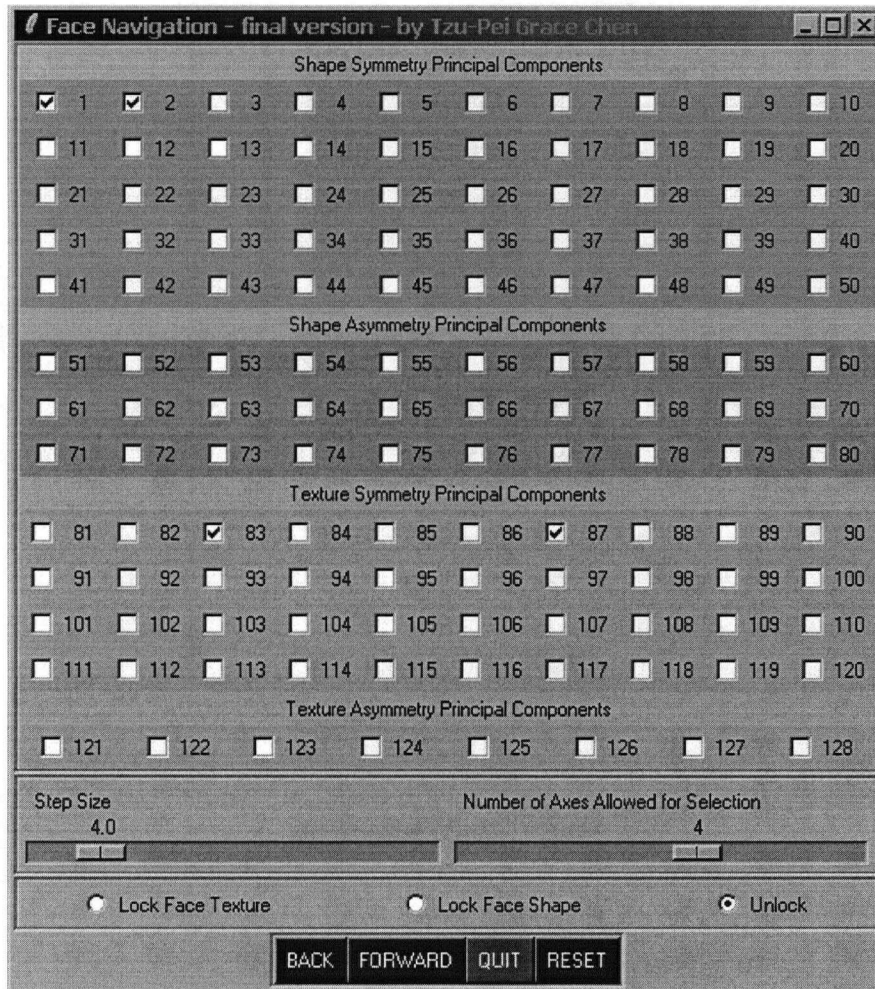
Figure 4.5: The control panel of our system allows users to manually select up to six of the 128 principal components and adjust various settings. The navigation axes are displayed in the order they are selected (anticlock-wise). This allows users to navigate in the *FaceGen* space. Notice there is only one slider controlling the step size for all navigation axes. This is a minor limitation of the current version and can be improved by having six individual step size sliders for each navigation axis.
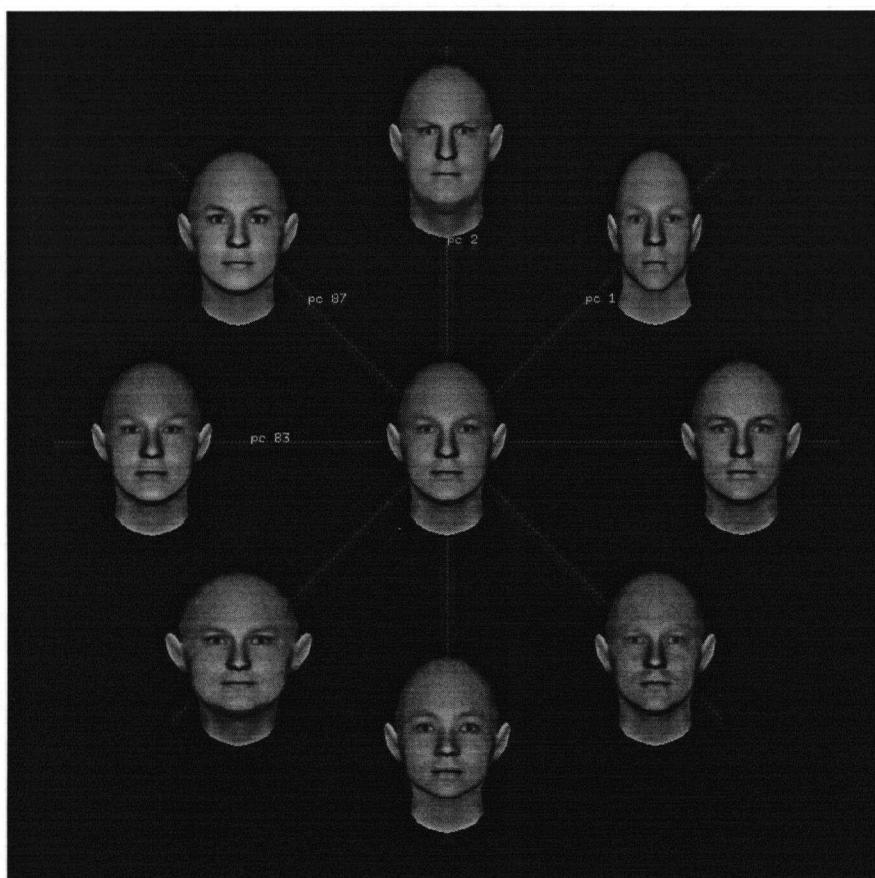
Figure 4.6: The settings on the control panel in Figure 4.5 determine the faces displayed here. The axes are labelled by their principal component numbers.

The Colour Metaphor

We design our interface by adopting colour as a metaphor. Colours have been studied extensively in computer graphics, and there are many system presentations of colours [52]. Although the number of facial dimensions has not been fully quantified, our concept is to think of face dimensions similarly to colour dimensions. If showing a colour spectrum helps users match colours, then it is likely useful to apply a similar interface to face matching. (In fact, a recent study by Yip et al. [67] has shown that colour does contribute to face recognition.) Our approach is similar to Adobe Photo Deluxe's interface for modifying colours of images.

The program presents users with a hexagon, initially made up of a user's picture in the centre and six neighbouring pictures (see Figure 4.7). The neighbouring pictures are essentially created by the original picture with a small amount of primary (red, green, and blue) and secondary (yellow, cyan, and magenta) colour variations. A selection of one of the neighbours moves the entire hexagon along that direction with the updated current centre being the previous neighbour selected. In this way, selecting the yellow neighbour makes the overall group of pictures one degree more yellow. On top of the hexagon, the original picture is juxtaposed with the current central picture of the hexagon so users can perceive the changes clearly.
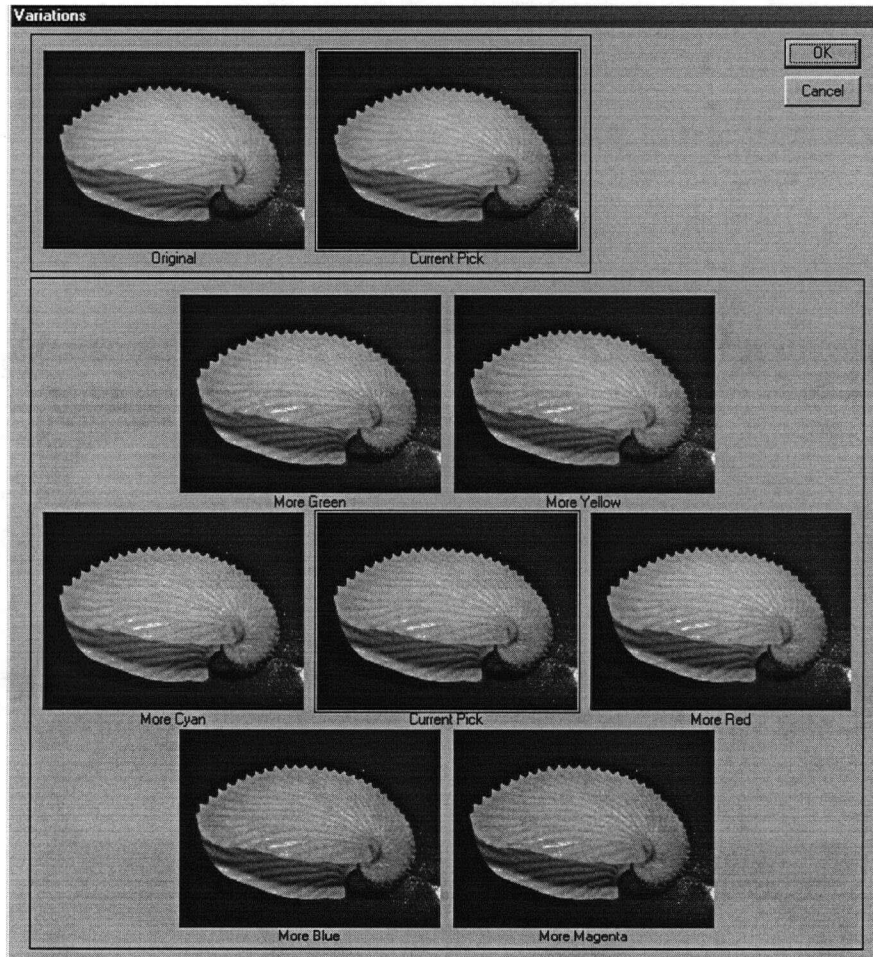
Figure 4.7: A snap shot of the Adobe Photo Deluxe interface. The original picture is a white shell with bluish background. Since no modification has been done to the image, the "Current Pick" image looks like the original as show by the top row shell images. The hexagon formed by seven shell images shows different colouring of the shell with small amounts of primary and secondary colours.

What is interesting about this approach is that movements in the 2-D screen space move along possibly correlated dimensions of colour space. Although the colour dimensions have labels, labelling may not be essential because users can see from the picture itself the changes they are making in colour adjustment. They can also see the gradient of colour change along any particular direction in the screen space. Thus, for attributes that do not correspond to easily identifiable axes, this approach provides the information the users need without complicating the issue with an interpretation of any particular dimension. This approach seems well suited for configural face space as we suspect that feature labels for dimensions are not useful, as discussed in the previous chapter.

Applying the same idea to our interface, we allow users to navigate from the average face to a distinctive face. Figure 4.8 depicts the skeleton for three navigation axes. If the current face can be parameterized as (x, y, z) which not only indicates its attributes along each axis, but also its location in the face space, then its neighbours would be those faces that are one step away from the current position along each of the x, y and z axes.

Figure 4.9 further articulates this idea and links together Figure 4.8 and Figure 4.1. Users see the portion of face space enclosed in the hexagon. Selecting a node within the hexagon repositions the hexagon window so it is centred on the selection. Selecting a node outside the hexagon is not feasible as it is outside the users' viewing. In its simplest form, only the one step neighbours are displayed but with the "zoom" function, the hexagon can increase its size to include more faces for users' viewing. This effects how many steps users can move at a time.
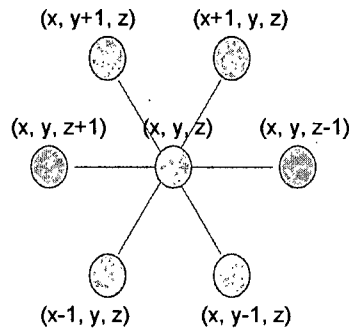
Figure 4.8: If the current face is located at (x, y, z), the neighbouring faces are those one step away along the x, y and z axes.

If users can only see the one-step neighbours, they can only move one step at a time. If they can see the one-step and two-step neighbours, then they can move up to two steps at a time.

Figure 4.10 depicts the interface for three navigation axes. The central face is the average face and its face value coordinates are $(0, 0, 0)$. Figure 4.11 shows the result after a user selects the face on the right. All the faces now take on some quality of the current central face. Note that this does not occur with the face on the left as it is the average face. However, the average face is precisely the result when you add two opposite faces together. The face and its anti-face cancel each other out by giving you the average face. Navigation in this face space proceeds by users selecting the face on the wheel that looks closer to the target face.

Note that the cancelling of a face and its anti-face does not translate to colours. That is, a colour and its complementary are not the negative values of each other. In fact, there are really no negative-value colours, and most colour
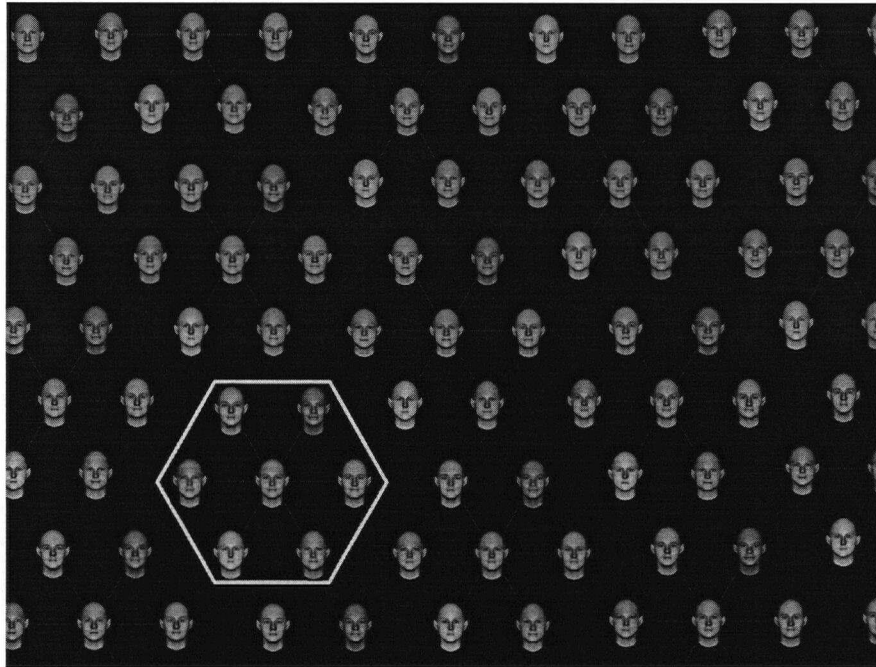
Figure 4.9: This image illustrates a conceptual view of the face space and the navigation interface. The hexagon window symbolizes the wheel interface. It acts like a sliding window showing users a part of the face space at every navigation step. Users can move this window by selecting faces at the edge within the hexagon.

systems do not allow users to select colours by a complementary scheme. The most one could get from navigating toward the "negative" direction of a colour is the darker hue of that colour, not its complementary colour.

It may also be evident that our interface resembles a pie menu [26, 55]. While our interface and pie menus are similar in geometry, our interface is much closer to a sliding window. Unlike pie menus, the centre of our wheel interface changes frequently during navigation. In other words, clicking on one of the neighbouring faces does not evoke a sub-menu to be displayed adjacent
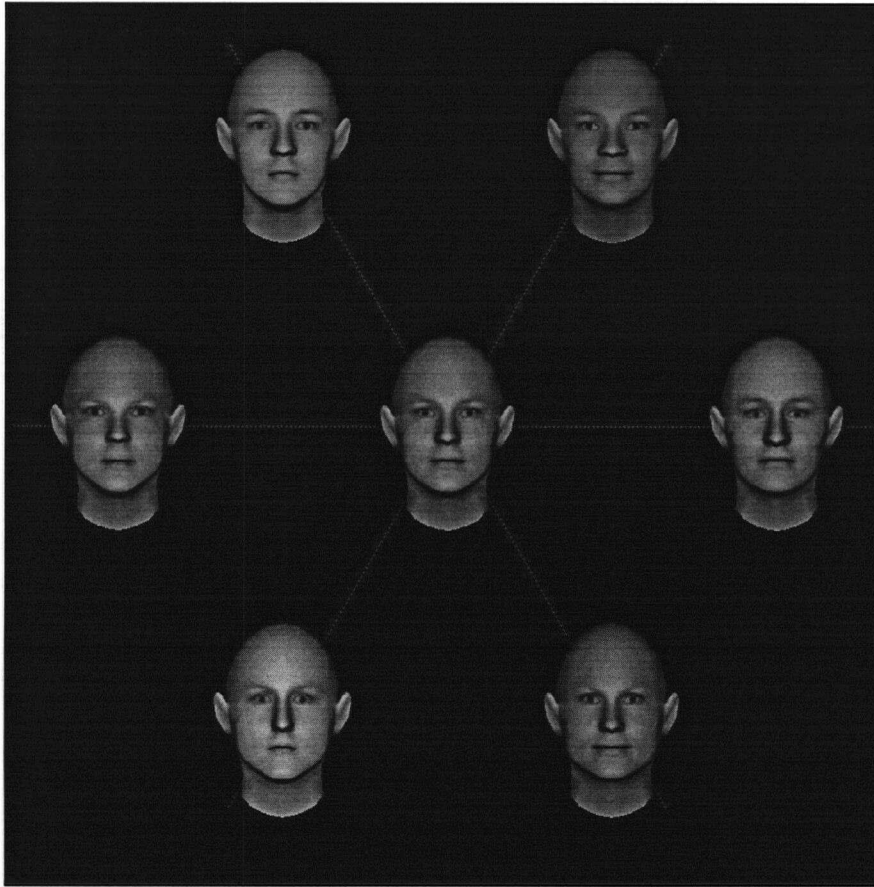
Figure 4.10: Configuration at the origin.

to the current menu. Instead, the entire display becomes the sub-menu. If we use the sub-menu approach, sub-menus and current menus will clutter together, forcing our design to display faces of a smaller size. This simply does not work as we notice that different face sizes make comparison difficult, and can therefore, impair users' navigation.
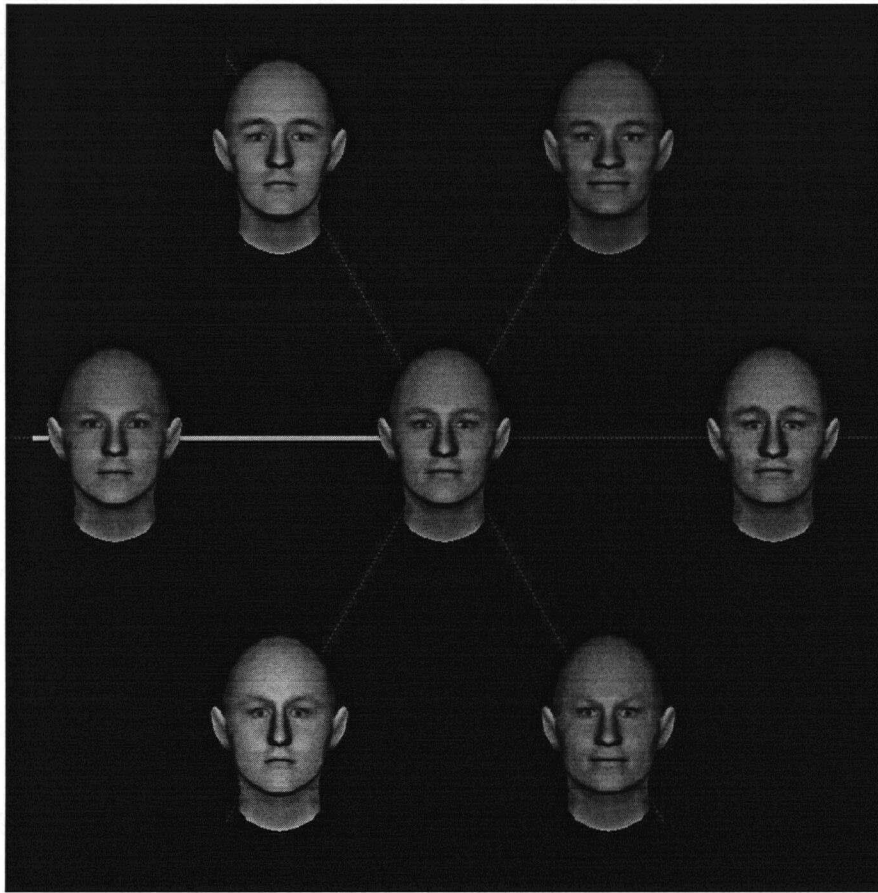
Figure 4.11: Configuration after moving one step to the right. Notice the gray line points to the average face signalling the direction for "undo."

Zooming Function

In addition to the "colour-wheel" interface described above, we investigate additional features to allow users to change the navigation step size. This allows users to zoom-in and out as necessary, similar to a map, and helps users to see the kinds of faces they can expect if they navigate along a particular axis at a particular step size.

We realize that we used little of the map metaphor. Initially, we were going to implement a navigation map using the Design Galleries idea [37] as it offers insightful visualization techniques. However, this concise bird's eye view of the face space gave way to our desire to emphasize the navigation feel. Moreover, it is difficult to determine how to arrange faces on a map meaningfully. One could associate faces by countries and display them via a world map. However, different geographical regions cannot detect certain unique looks. For example, Bush-men in Africa can appear to be rather Oriental. The map metaphor also restricts dimensionality to two, further constraining the number of navigation axes we would like to show. Hence, this idea did not go very far.

Additional effort is given to explore the possibility of a three-dimensional face map. We have limited success with this because of lack of ease of navigation, it is probably better to use a three degree of freedom input device (a mouse only provides two). Different viewing angles are also essential to accommodate the multi-dimensional nature of face space. Although navigating in 3-D face space is a very compelling experience, the limited computational power of our machine disturbs the fluidity of displaying and animating a large number of faces. Avoiding this issue then prompts for design alternatives which restrict the number of faces shown at every navigation step. This involves consideration of what sample of the face population to display and how to prioritize faces accordingly. For example, one can consider which plane of the face space to render. One can also use opacity of faces as visual cues to indicate order of importance. Making all faces face the user also facilitates comparison. All these

considerations, although sensible, complicate the interface design. Ultimately, the three-dimensional face map idea is not flexible enough for our needs, and that is why the radial axes colour scheme is finally chosen.

"Back" Function

We add a "back" mechanism similar to a web browser so that users have a thread tying them back to the origin if they are ever lost in the face space. The "back" direction is indicated by a thick gray line (see Figure 4.11) which points to the direction one should navigate to "undo" the last movement.

This is a beneficial and important feature as users can retrace their steps to see what faces they have already visited. As Baker [5] suggests, a retrace mechanism can counteract users' mental image degradation since they can review their paths. The concept of "direct manipulation"[1] by Shneiderman [53] also listed that reversible commands should be an essential component of good interface design. We store the face locations users visit in a history stack (see Figure 4.4), and the stack entries form the navigation path.

Face Management

When evaluating the user interface, it is important that new faces are displayed quickly so that users do not have to wait too long for the display to update whenever they click the mouse. We implement background face loading mecha-

---

[1]According to Shneiderman's definition, a direct manipulation interface should: present the objects visually with an understandable metaphor; have rapid, complementary and reversible commands; display the results of an action immediately; and replace writing with pointing and selecting.

nisms to improve drawing efficiency so that users are more engaged in the flow

of interaction.

This is how it is accomplished. When users navigate to a new location in face

space, the program generates a list of its first round neighbours and several lists

of its second round neighbours[2]. The rendering function checks if the faces of the

first round neighbours, that need to be displayed immediately, already exist in

the database[3]. If not, it requests the face generator to create and load these into

memory by using the OpenGL display list. If the faces are already registered in

the database, then their display list indices are returned to the calling function.

Only when the rendering function is complete does the background face loading

take place. The background face loading is performed by the OpenGL idle

function. It loads the second round neighbours of the current face one at a time

and can be interrupted whenever the rendering function is activated. In this

way, like a cache with pre-fetching mechanism, only the faces that need to be

seen are generated and kept in memory.

This is helpful in engaging users in the navigation process because as users

navigate, it may take them some time to view the faces. While the interface

remains idle, faces are loaded in the background so that when users make a

selection, faces appear to load quickly, providing the illusion that the face space

already exists.

---

[2]The second round neighbours are neighbours of neighbours. If there are twelve first round neighbours, then there are at most 144 second round neighbours. Note that some of the neighbours are repetitive and our program checks if the neighbouring faces already exist in the database before creating them.

[3]Our face database is made of a multi-dimensional list structure. The standard C++ library provides an efficient list structure, and we have not found the need to use other data structures such as a heap.

Locking Mechanism

In order to facilitate face navigation, we implement a locking mechanism that allows users to lock the face shape or texture. This allows users to isolate the shape or texture of faces in order to simplify the face searching tasks. This function is accessible via radio buttons in the control panel (Figure 4.5).

The Control Panel

The wheel interface is complemented by a control panel (Figure 4.5) which allows all 128 principal components to be accessible to users. It also includes controls of other variables, such as the number of axes to display, the step size and other useful functions. The control panel communicates with the wheel interface window via a socket. Our program deciphers the message sent by the control panel and updates the variables accordingly (see Figure 4.4). Although we have not found a way to organize and label the principal components meaningfully to facilitate users in selecting desirable navigation axes, having these controls allows us to obtain a good sense of the *FaceGen* face space. Consequently, we have a good sense of the range of face parameters for devising our finite and holistic face spaces.

Notice that one of the restrictions of the control panel is the one slider control for the step size for all navigation axes. The initial design has sliders for all 128 principal components, but such an interface can be rather overwhelming. Since it does not impede the exploration of the face space, this mechanism is not modified. Once we select the principal components we want to use for our

experiments, we can adjust the step size individually for the navigation interface tailored for the experiment. A dynamic set of up to six sliders for the principal components selected would have been ideal.

We have thus far described the design decisions that influence our general face navigation system. The system architecture is also laid out with each component expressed. In order to test our interface, however, customization is required to tailor to the purpose of the experiments. We describe subsequent development in the next section.

4.3 Subsequent Development for Testing

A useful approach to test our wheel interface is to compare it with a conventional interface. The interfaces compared must be given similar contexts and equal advantages as much as possible for fair judgement. We describe our challenger interface below and additional features developed as part of the testing interfaces. Customizations specific to each experiment are not included here but contained in later chapters conveying each experiment.

4.3.1 The Challenger: Sliders

An obvious approach for navigating the same face space described above is to use a slider since it is a common interface for multidimensional parameter manipulation such as used for colour selection. In order to observe people's preferences, we implement a slider-like interface comparable to our wheel interface for face navigation. These are not true sliders because we discretize our face space;

therefore, our "sliders" are not continuous. They are rather similar to a string of radio buttons. Due to computational speed limitations, our sliders also do not activate updates of faces until users release the mouse buttons. Thus, users cannot slide around to gain a quick glimpse of faces. Despite these limitations, we explore two different ways to implement our sliders to provide a continuum of fair alternatives to the wheel. Through this, the discrete nature of the sliders does not limit our experiment.

Our first approach, called the *static sliders*, is the most straightforward. However, for equal means of comparison, so that users have the same information on the screen as on the wheel interface, we develop a modified slider approach called *dynamic sliders*. Figures 6.3, 7.5, 7.3, 8.3 and 8.5 show snap shots of the slider interfaces used in our experiments. Note that they have a face line-up at the bottom of the window. This is to facilitate face matching, and it is described in a later section.

The Static Sliders

The sliders are discretized to provide m number of possible values each depending on the fineness of resolution. Our database is then of the size m^6 number of faces. Our studies use two values of m, namely 5 and 9. Here we describe the case with $m = 5$ in more detail. Users can move the slider position back and forth along the five possible values. The discrete values of -2 to 2 will show above the sliders to indicate the value of the current slider ball position. The wheel interface is limited in that it only allows users to move one step at a

time even though the version with zooming does allow users to adjust the step size. The sliders also have a "snapping" mechanism in which the sliding ball is always snapped to the closest discrete position when users release the mouse. This reinforces the discrete options. If users prefer, they can also click on the end triangles to move the slider position one step at a time. This accommodates those users that prefer to click rather than drag sliders.

Either end of a slider shows the face that users can expect at the extreme of that navigation axis with respect to the origin. This information is given so that slider control is meaningful to users. The task for the user is to adjust the sliders by trying to find a combination of qualities presented by the extreme point faces. These faces, however, remain fixed during the navigation process, hence, we call this interface *static sliders*. To make it even clearer, that is, if the current position in face space is (x, y, z), the extreme faces shown for the first static slider is $(-e, 0, 0)$ and $(e, 0, 0)$, where e is the extreme value. Thus, the extreme faces are not affected by the current position.

The Dynamic Sliders

We modify the slider interface to dynamically update the end point faces of the sliders. The reason this may be important is that changing one slider value alters the effect that the others would have if changed. Thus, after a slider is changed we can update what the end points of each slider would look like. Either end of a slider therefore shows faces that users can expect if the sliding ball is positioned at the extreme points. In other words, if the current position in the

	Wheel	Static Sliders	Dynamic Sliders
Movements	One step at a time	multiple steps	multiple steps
Faces Shown	Nearest neighbours of current	Extreme neighbours of origin	Extreme neighbours of current
Face Update	Dynamic	Fixed	Dynamic
Extra	"back" function for undo	None	None

Table 4.1: Features of the navigation interfaces used in the experiments.

face space is (x, y, z), then the extreme faces shown in the dynamic slider of the first axis are $(-e, y, z)$ and (e, y, z). This update gives users particularly good visual cues if they reached the neighbourhood of the target face. The problem with dynamic sliders though, is that all the displayed faces are changing with every selection possibly confusing the users. They may not be able to tell what any particular slider does as the end points keep changing.

4.3.2 Additional Features

In order to compare our wheel and slider interface design, we modify our interfaces for different conditions of user testing. Features of each interface used in our studies are summarized in Table 4.1. Although our wheel interface has many more features than sliders, we only list the ones used for the experiments since we stripped down the wheel navigation interface for fair comparison with the sliders. What is important to know is that in giving the sliders more of an advantage, we do not make the zoom function of the wheel interface accessible to subjects. The locking mechanism is also excluded as the face space is not large enough to render it useful. However, the "Back" function is included.

There are a number of new features. For those that are consistent in all

experiments, we describe them in this section. For those customizations that are specific to each experiment, we describe them in details under the "Apparatus" sections of the chapters on experiments.

The new features are as follows:

- Face Line-up.

- Face Space Boundary Indicator.

- Memory Checking Mechanism.

- Control Panels.

- Face Targets.

- Navigation Axes.

The top three items are described in this section. A face line-up is included for both wheel and slider interfaces to facilitate face comparison. For the wheel interface, we provide visual cues to inform subjects when they have reached the edge of the face space—hence the boundary indicator. There is also a memory check mechanism to ensure the program runs stably.

The last three items are described in later chapters. Control panels are created to accompany each experiment to allow users to toggle between testing and practise mode. Face targets are selected in particular ways for each test. Lastly, the navigation axes, formed by principal components, determines the appearance of the faces navigated.

The Face Line-up

A common addition to all interfaces used for testing is the inclusion of a face line-up (Figure 4.12). This is for accommodating users in a face match via navigation in our face space. We line up the average face, the current face and the target face from left to right and position them at the bottom of the interface window. The reason that it is placed below is because priority is given to the part that accommodates face selection. The display of the average face (labelled as "original") serves as an anchor so that users are aware that they are navigating away or toward the origin when exploring. The target face and the current face are juxtaposed for ease of comparison. The current face is also shown in the centre of the wheel interface.
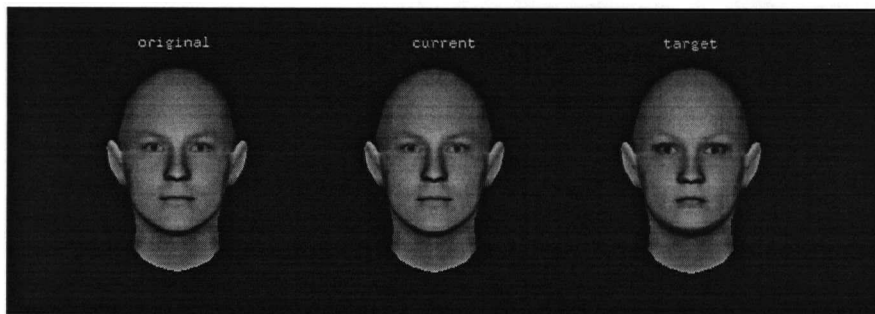
Figure 4.12: To facilitate face matching tasks, the average face (labelled "original"), the current face and the target face are displayed from left to right at the bottom of the interface window. Notice the current face is the original average face at the beginning of navigation.

When users navigate to the target destination, a halo is placed on top of the current face with a message above it that notifies the users of success (Figure 4.13). During the practise mode, the message prompts users to select a different

condition to practise. During testing mode, on the other hand, the message instructs the subjects to press the space bar to proceed to the next trial.

Figure 4.13: When the face target is successfully matched, a message is displayed above the face line-up. The current face, which now looks like the target face, has a halo on top.

The Boundary Indicator

When users reach the edge of the face space in the wheel interface, a rotated wire cone is displayed in place of the face to indicate that users cannot navigate any further (Figure 4.14). The slider interfaces do not need this feature as the width of the sliders indicates how far users can go.

Memory Check

For stabilizing the performance of our program, we include a checking mechanism for the face database so that our program does not consume all the memory. Once the number of faces loaded into the memory exceeds 100, unnecessary faces can be discarded. The discarding algorithm looks for the face furthest away from the current face that does not need to be displayed immediately. Such a face
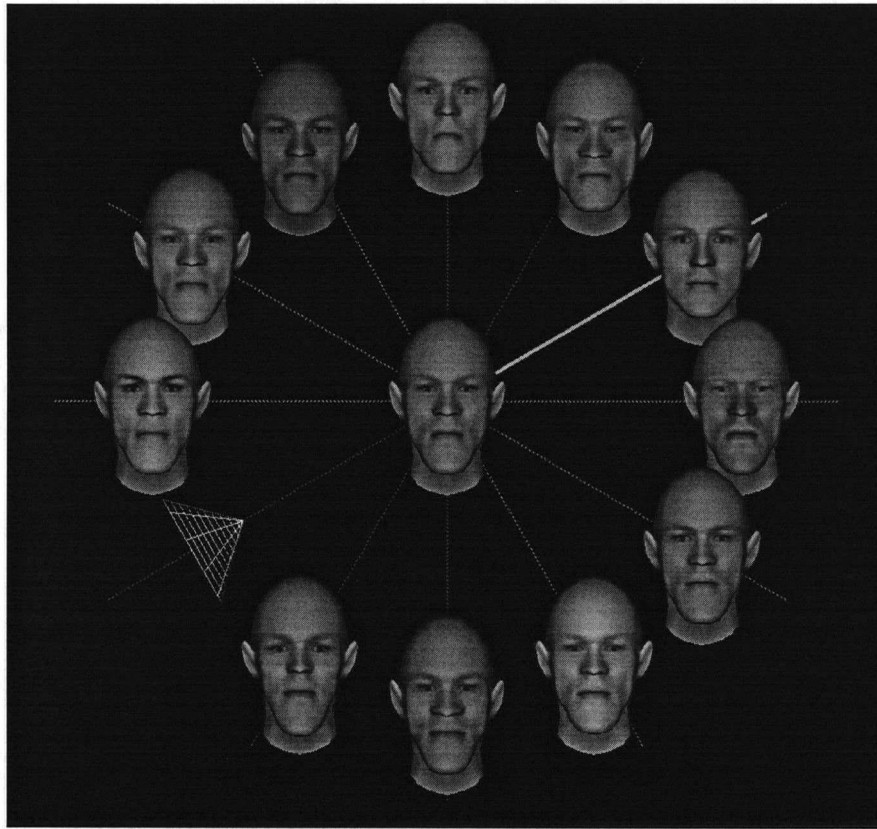
Figure 4.14: When the edge of the face space is reached, a wire cone is shown in place of the face.

can be replaced by a new face as users navigate.

We have, in this chapter, provided reasons why our face navigation system takes on a configural approach with the use of eigenface. Reliance on users' face recognition ability is a key feature in our navigation system. We describe our system architecture. Starting with the face space, we define its structure and explain how it is created. Then the wheel interface is described which is our main creation for face navigation. We outline subsequent features developed for

the purpose of user testing. The challenger interfaces are two slider-like interfaces which we create for comparison with the wheel interface. Other features include the face line-up, the visual indicator of face space boundary and memory checking mechanism. Three more specialized customizations: the control panels, the face targets and navigation axes will be revealed in chapter 6, 7 and 8.

Before further details on the interface customization are given, a user testing overview is presented next.

Chapter 5

User Testing Overview

The previous chapter outlines our general system architecture. For user testing, however, our interface as well as the challenger interface need to be modified in order to isolate the factors we wish to investigate. In this chapter, we provide the skeleton of our experiments, so that readers have a comprehensive view of all the testings done.

We lay out each experiment in the first section with emphasis on the independent variables explored. The independent variables are essentially the factors believed to have an effect on the outcomes. They help quantify the input stimuli. Following that, we turn our attention to the experimental design. We explain how and why we choose a factorial within-subject approach. We then give a brief description of how we plan to analyze the data of the dependent variables. The dependent variables are inherently the measurements derived from the data taken during the experiments. They are used in the analysis that weighs the significance of the factors—particularly the independent variable, interface.

5.1 Experimental Layout

We carried out four experiments in total following examples in [43], [50] and [39]. A number of independent variables are chosen for investigation. We mention

Test	Interface	Axes Type	Number of Axes	Target	Resolution
1	static slider vs. wheel	coupled	3 vs. 6	N/A	coarse
2	dynamic slider vs. wheel	uncorrelated vs. correlated	6	N/A	coarse
3 & 4	dynamic slider vs. wheel	uncorrelated	6	near vs. far	fine vs. coarse

Table 5.1: Summary of the independent variables investigated in the three pilot tests and the formal experiment (test 4). Where more than one condition is given in a cell, it indicates the different treatments of the independent variable concerned. Otherwise, one condition indicates the variable is fixed in the experiment.

them in the subsections below and include the summary of the experimental conditions in Table 5.1. The first row of the table lists the independent variables.

5.1.1 Pilot 1

In this experiment, we are interested in finding out how well the wheel interface performs versus the static slider interface. We are also interested in varying the number of navigation axes to either three or six to see how this effects subjects' navigation. The independent variables of interest are therefore interface and the number of navigation axes. The other factors such as axes type and resolution remain fixed.

5.1.2 Pilot 2

For Pilot 2, the two independent variables of interest are the axes type and the interface type.

Since different kinds of navigation axes may evoke different reactions from the subjects, we are interested in exploring ways to combine principal compo-

nents as navigation axes in this experiment. We have three kinds of axes. Independent axes are basically uncorrelated principal components[1]; coupled axes are essentially positively correlated principal components; and correlated axes include both positively correlated and negatively correlated principal components. In this test, however, we only use correlated and uncorrelated axes since coupled axes are already tested in Pilot 1.

As for interface, we compare the wheel interface with the dynamic slider interface in this experiment. Static sliders are not used because we find them limiting. While fixing the number of axes at six, we test if one interface is better than the other for the different kinds of navigation axes.

5.1.3 Pilot 3 and the Final Experiment

In Pilot 3, we want to investigate if resolution and target type have any effects. We again use the wheel interface and the dynamic sliders. In this way, interface, resolution and target type are the three independent variables. We fix the number of navigation axes at six and use the uncorrelated type. Since Pilot 3 shows interesting results, the final experiment is set to be just like Pilot 3 except with a larger pool of subjects.

In previous experiments, the resolution is fixed at coarse level. That is, there are only five possible discrete positions along each navigation axis, providing a

---

[1]We are using the words "independent" and "uncorrelated" very loosely here. We refer to our independent or uncorrelated navigation axes to be the axes that allows separate manipulation of shape and texture principal components. These principal components are orthogonal in their respective shape and texture space, but may be correlated in reality. For example, imagine a data set of where all dark-skinned individuals have big chins. In this case, the texture and shape components are clearly correlated.

face space of 5^6. In Pilot 3 and the final experiment, however, we introduce a fine resolution which allows nine discrete positions along each navigation axis, providing a discrete face space of size 9^6.

We also explore the target type. Targets can be far or close to the average face at the origin. We want to investigate whether this has an effect on face navigation performance. Presumably, the further the target, the harder it is to reach.

Having established the independent variables in each experiment, we explain our experimental design next.

5.2 Experimental Design

All the experiments we conduct are within-subject factorial design requiring subjects to do face matching tasks by way of navigating in face space. The choice of a factorial design is due to a number of associated benefits. First, factorial-designed experiments are more efficient than one-factor-at-a-time experiments. Second, a factorial design is necessary when interactions may be present between variables. Such awareness of the interaction helps to avoid misleading conclusions. Additionally, a factorial design allows the effects of a factor to be estimated at several levels of the other factors, yielding conclusions that are valid over a range of experimental conditions [43]. Table 5.2 shows a summary of our factorial designs.

As for the reasons behind selecting within-subject experiment, the aim is to

Test	Independent Variables	Factorial Design Dimensions
1	interface, number of axes	2×2
2	interface, axes type	2×2
3 & 4	interface, resolution, target distance	$2 \times 2 \times 2$

Table 5.2: Summary of the factorial designs of three pilot tests and the formal experiment (test 4). The dimension of the design is determined by the number of variables and the number of treatment levels per variable. For example, two independent variables with two levels each gives a factorial design of 2×2.

minimize variations due to individual differences [39]. Such design allows collection of data from fewer subjects but has the disadvantage of raising potential fatigue and practise effects. This is avoided in our experiments by randomly determining the order of the testing conditions so that the "order effect" is cancelled out [50].

5.3 Analysis Plan

Application of factorial design calls for analysis of variance (ANOVA). Although ANOVA is a powerful technique which allows one to test more than one factor at a time, we are mainly interested in testing the interface variable. We want to know if the wheel interface is better than the two slider interfaces; more importantly, under what conditions is one interface better than the other.

In order to determine what factors are significant, ANOVA test is applied to a set of measurements (dependent variables). They are numbers that we deduce from the data. Through ANOVA statistical calculations, we can then see whether we accept or reject our null hypotheses. We describe the null hypotheses in the next section then the dependent measurements afterward.

5.3.1 Null Hypotheses

Our null hypotheses are listed below. Since we are mainly testing for the mean difference of the data from the different interfaces, the hypotheses are geared toward the interface variable. The null hypotheses are different from each other in their dependent measures only.

- H_{01}: Subjects take a similar amount of time using the wheel interface and the slider interface.

- H_{02}: Subjects score similarly using the wheel interface and the slider interface.

- H_{03}: Subjects revisit faces a similar number of times using the wheel interface and the slider interface.

- H_{04}: Subjects see a similar number of unique faces using the wheel and the slider interfaces.

- H_{05}: Subjects visit a similar number of faces during the "refinement phase" of face matching tasks, using both the wheel interface and the slider interface.

- H_{06}: Subjects visit a similar number of faces during the "approaching phase" of face matching, using both the wheel interface and the slider interfaces.

Naturally, ANOVA can also indicate variables other than the interface factor to be significant. We are not as interested in testing the null hypotheses of those;

Test	Dependent Variables
1	time, score, repetition, uniqueness
2	time, score, repetition, refinement efficiency
3 & 4	time, score, repetition, refinement efficiency, approaching efficiency

Table 5.3: Dependent variables used for the three pilot tests and the final experiment (test 4).

hence we exclude them here.

5.3.2 Measurements

As we pointed out before, the hypotheses differ mainly in dependent measurements. Our dependent measurements or dependent variables are time[2], accuracy score[3], repetition[4], uniqueness[5], refinement efficiency[6] and approaching efficiency[7]. We do not test all hypotheses in each experiment because some dependent variables do not apply. Table 5.3 summarizes the dependent variables used in each experiment. This table recalls Table 5.2 which summarizes the independent variables and factorial design dimensions for each experiment.

We have briefly described the dependent measurements in the footnotes but

---

[2]Time is measured in seconds per trial.

[3]Score is assigned in such a way that if the final target takes, at most, n number of steps and the subject only manages to come to its immediate neighbourhood, i.e. one step away from the target, then her score is $\frac{n-1}{n}\%$. If the subject is k steps away, then his score is $\frac{n-k}{n}\%$. Note that step is the step in the discretized face space we devise for user testing, not the step size of *FaceGen* face space.

[4]The repetitiveness measure is determined by the number of times subjects revisit locations in the face space.

[5]Uniqueness is measured by the number of unique faces subjects encounter per trial.

[6]Refinement efficiency is measured by counting the number of face locations in face space subjects visited that fall within two steps from the face target.

[7]Approaching efficiency is measured in terms of the number of face locations visited by subjects that are more than two steps away from the face targets.

two require special attention. The idea of exploring the refinement and approaching stage of navigation is borrowed from Schwarz and further clarification is necessary.

Approaching Phase versus Refining Phase

There are some subtle differences between approaching and refining. By "approaching," we mean "finding the neighbourhood" of the target face. By "refining," we mean the navigator has arrived in the neighbourhood and is in the process of reaching the target. Schwarz [52] uses the term "converge" instead of "approach," which is inappropriate as "converge" includes reaching the neighbourhood as well as meeting the targets after some exploration. We differentiate between the stage of finding close proximity and the stage of refinement for reaching the target.

It is, unfortunately, not straightforward to pinpoint a period where users approach the neighbourhood of the target and a period where users refine a match. This is because they can move between these stages before reaching the goal. Nevertheless, we seek a way to break down the data. We categorize the two-step neighbourhood of the target to be the "refinement region," and anything outside of that to be the "approaching region." In this way, when we refer to "refinement phase," we mean users have reached the two-step neighbourhood. Similarly, when we use "approaching phase," we mean users are still looking for the two-step neighbourhood. Figure 5.1 illustrates the point. What we are essentially doing is simply splitting the face locations subjects visited

into the refinement and approaching categories.
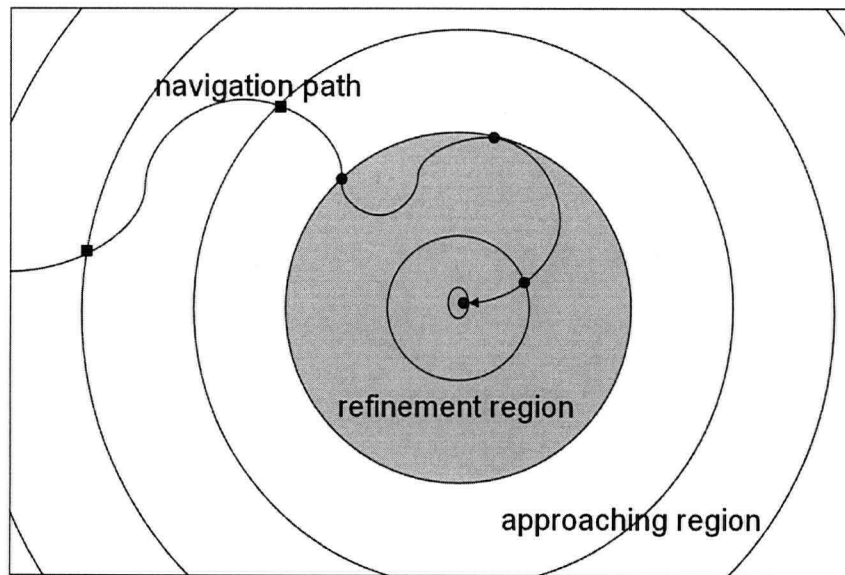
Figure 5.1: The image illustrates the conceptual idea of refinement versus approaching. The centre of all circles is an oval which indicates the target face. The wavy arrow, which points to the target face, is a navigation path. The region highlighted in cyan, i.e. the area enclosed by the second circle around the oval, is the "refinement region." If subjects navigate to that region, they are "refining" their face match to get to the target. However, if they are outside of it, they are considered to be "approaching" the two-step neighbourhood. Each circle represents a spatial distance (in user space steps) from the target. The small red circles symbolizes the face locations the navigator visited which count toward refinement. The small blue squares indicate the face locations the navigator visited which count toward approaching.

At this stage, readers may wonder why we select two steps to be the size of

the neighbourhood in our "refinement phase" definition. We choose two because

it seems to be the threshold where differences in our data are observed. It is

debatable how large the refinement neighbourhood should be, but we feel two

is a good number to start with.

We have presented and explained all the dependent measurements which are needed for ANOVA test to determine which independent variables are significant. We introduce additional statistic tests in the next section.

5.3.3 Follow-up Statistical Tests

As with many data analyses, often one statistical test is not enough. Follow-up tests may be necessary in search of a convincing explanation. Note that according to Roberts et al. [50], post-hoc tests following ANOVA (such as T-test or Tukey test) is only necessary if there are independent variables with more than two levels. Since we only have at most two levels for each variable, post-hoc tests do not apply to our results. However, for ANOVA tests that show significance in interactions, "simple main effect analysis" is applied to determine which interaction condition is significant.

We have provided an overview of our experiments in this chapter in an attempt to set the stage for the experiment discourse to come. We lay out our experiments, emphasizing the independent variables in each. We explain our experimental design, and outline the plan for data analysis by describing how the null hypotheses and dependent measurements fit together. We pay close attention to our definitions of "approaching" and "refining" phase as we categorize our data in special ways. In addition, we touch on the follow-up tests to ANOVA to make readers aware of other statistical techniques employed.

The main concept we try to convey in this chapter is as follows. We adjust

the stimuli subjects experience through varying the levels of treatment (i.e. the values for the independent variables). On the other hand, subjects' interaction with our interfaces offer the data which essentially form the values for our dependent variables. In other words, independent variables are the input while the dependent variables are the output of the experiments.

By providing both foci on independent and dependent variables, we thus paint a picture of the basic ins and outs of user testing. We have also pointed out that our main desire is to see if the interface factor is significant, for it is in exploring and understanding the strengths and weaknesses of both interfaces that we make a contribution to research.

Having established the skeletal structure of the experiments, the body of the three pilot tests and the formal experiment are further articulated in the next few chapters.

Chapter 6

Pilot Test 1: Effect of Interface and Number of Navigation Axes

6.1 Introduction

In our first pilot study, we are interested in comparing the wheel interface with the static slider interface under two navigation axes conditions. The first condition provides the subjects with three navigation axes and the second condition provides the subjects with six navigation axes. The two-variable factorial design is summarized in Table 6.1 showing two levels for each of the independent variables.

We expect the difference between the wheel interface and the static sliders to be large because the former provides more information than the latter. We also believe if the number of navigation axes increases, the difficulty in face matching increases as well.

	Number of Navigation Axes	
	Three	**Six**
Interface **Wheel**		
Static Slider		

Table 6.1: Two variable factorial design for Pilot 1.

6.2 Method

6.2.1 Subject

Six graduate students volunteered from Electrical Engineering as subjects for each experiment. Of the six subjects, one is a woman and five are men; and four are Caucasian, one is East Indian and one is Oriental.

6.2.2 Apparatus

Given the general features of the interfaces (outlined in Chapter 4 on implementation), we describe specific customization in this section. The modifications include the addition of a control panel, special selection of face targets and "coupled" navigation axes used.

The User Testing Control Panels

In order to make things simpler for subjects during user testing, we have specially designed control panels for our experiments. Unlike the control panel of our full version (Figure 4.5), the ones used for the experiments are much simpler for they are used to facilitate user testing procedures. Figure 6.1 shows the control panel with the four different testing conditions for Pilot 1. Subjects can pick any one of the conditions during the practise mode. However, when subjects select the "Start Test" radio button, all mechanisms are automatic from then on.
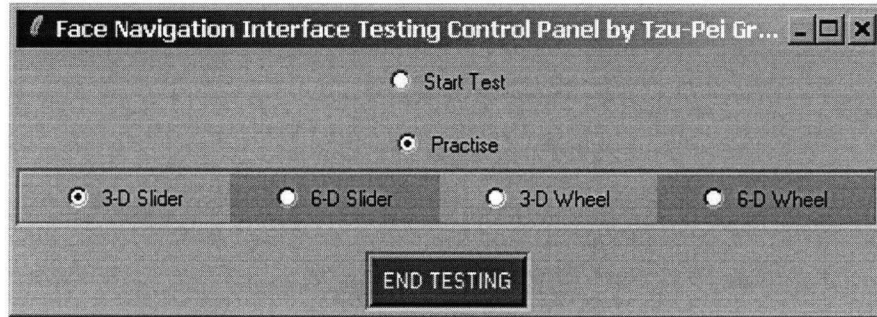
Figure 6.1: Control panel for Pilot 1. Note we use "3D" to indicate three navigation axes and "6D" to indicate six navigation axes.

Condition	Set of Numbers	Example Target
Three Dimensions	$\{1, 1, 2\}$	$\{1, 2, -1\}$
Six Dimensions	$\{1, 1, 1, 1, 2, 2\}$	$\{1, 2, -1, -1, 1, 2\}$

Table 6.2: In Pilot 1, targets are made up by the set of numbers in the second column. Each target face is any sequence of the numbers in the set, which can take up positive or negative signs indicating directions along the axes. The third column shows some examples.

The Face Target

Targets for experiments are specifically selected so that they all require the same number of steps over all navigation axes to reach, while maintaining the same spatial distance from the origin.

In Pilot 1, the target for three navigation axes test cases are chosen so that one of the axes requires two steps while the other two require one in either direction. Similarly, the target for the six navigation axes test cases are chosen so that two of the axes require two steps, and the rest require one step only (see Table 6.2 for a clearer symbolic representation).

Axes	Shape			Texture		
	PC	Range	Step Size	PC	Range	Step Size
A_1	S_{11}	$[-10; 10]$	5	T_1	$[-2; 2]$	1
A_2	S_{12}	$[-10; 10]$	5	T_2	$[-3; 3]$	1.5
A_3	S_{13}	$[-12; 12]$	6	T_3	$[-4, 4]$	2
A_4	S_{14}	$[-12; 12]$	6	T_4	$[-9; 9]$	4.5
A_5	S_{15}	$[-12; 12]$	6	T_5	$[-5, 5]$	2.5
A_6	S_{16}	$[-10; 10]$	5	T_6	$[-8; 8]$	4

Table 6.3: Coupling mapping of navigation axes in Pilot 1.

The Navigation Axes

The navigation axes determine the faces created and the structure of the face space.

In this first pilot test, the choice of the number of axes is either three or six, and each axis is made from a pair of principal components. Table 6.3 shows the coupling of the shape principal components and the texture principal components for our navigation axes. Using the first row as an example, navigating one step along axis A_1 is equivalent to moving one step of size five along principal component S_{11}, and one step of size one along T_1 within their realistic range.

Although the choice of principal components may appear arbitrary, we combine these principal components because they provide a large variety of faces that are convincing—such as faces that anyone may encounter on the streets. The selection is based on careful viewing of all face types along the 128 principal components via our full version. The assumption is to think of faces as colours. From Schwarz's work [52], we do not expect subjects to exhibit a large variation in colour navigation (be it via RGB or CMY) as long as the principal

colours are quite distinguishable. Similarly, by selecting higher level principal components where resulting faces are very distinguishable, subjects should be able to tell them apart. This is the best we can do for now without sophisticated psychological experiments which may extend beyond the scope of this thesis.

A realistic range is determined by manually adjusting the step size of the control panel of our full version to view the faces along the principal components of interest. Although *FaceGen* produces high quality 3-D morphed faces, it has its imperfections. Users can tell when they reach an unrealistic range when produced faces appear caricature-like and have unnatural streaks and skin tones which often occur around the eyebrow areas. This is due to the high variations of face scans used to generate the statistical face model. The face scans may not be aligned exactly when the principal components are computed; hence there are minor artifacts.

An alternative to using coupled principal components is to have each axis correspond to an *independent* texture or shape principal component. We use this in Pilot 2. The reason that we do not use independent texture and shape principal components in Pilot 1 is that we want to create a wider variety of faces. Moreover, for the case where we only have three axes in the interface, if we use independent principal component mappings, then we would have an unbalanced choice of either two shape principal components and one texture component or vice versa. By providing coupled principal components, we obtain a balance of the face space spanned, as well as a good variety of realistic faces within which to navigate.

We hand design how much one unit of movement on the axes corresponds to changes in the value of the coupled principal components used to create each face. Note that the step size for shape and textures does not have the same values. The larger the principal component, the more variation it captures, so we adjust the step size for each accordingly. We find the range of the principal components in which faces appear realistic (by using our full navigation system (Figures 4.5 and 4.6)), and divide that range into four to obtain the step size. Therefore, we allow five possible positions (including the neutral position) along each navigation axes, making this face space a size of 5^6. In this way, users navigate in a discrete realistic face zone, not a caricature zone.

Snap shots of the interfaces used with these axes show the variety of faces in Figure 6.2 and 6.3. We precompute faces to avoid the lag of face updates the wheel interface might encounter.

6.2.3 Procedures

All subjects go through three stages of our experiments: practise, testing and interview. During the practise period, subjects are introduced to the interfaces and given time to practise. During the testing period, subjects complete two replicates of all possible testing conditions. Finally, during the interview period, subjects give us feedback of their thoughts from the testing experience. Appendix A shows sample questions of the interview.

We create a testing ground comprised of a six-dimensional face space. As mentioned before, the face space contains six axes with five values each, totalling
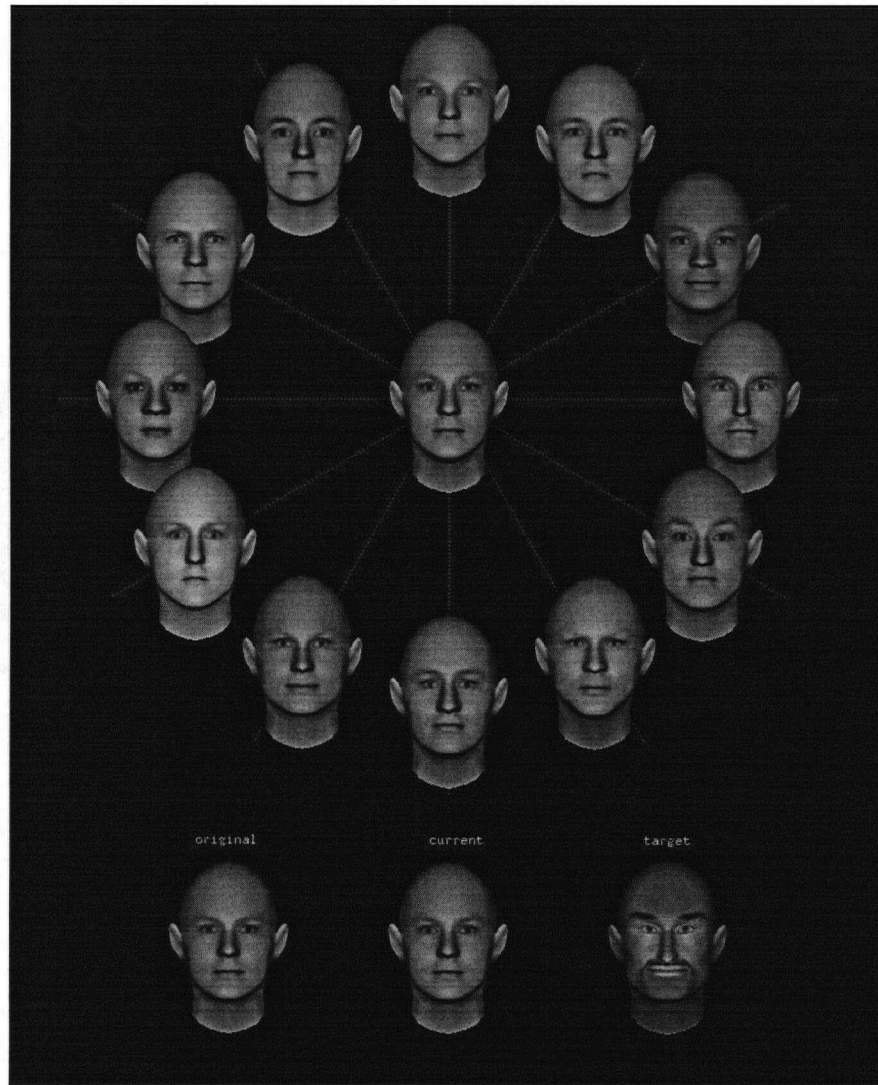
Figure 6.2: Six axes wheel interface used in Pilot 1. The user is required to try to find the target face (shown in the lower right corner) by clicking on the perimeter faces in the wheel. The current centre of the wheel is also shown next to the target face for easy comparison. The original (average) face is shown in the lower left corner to indicate the face at the starting point.
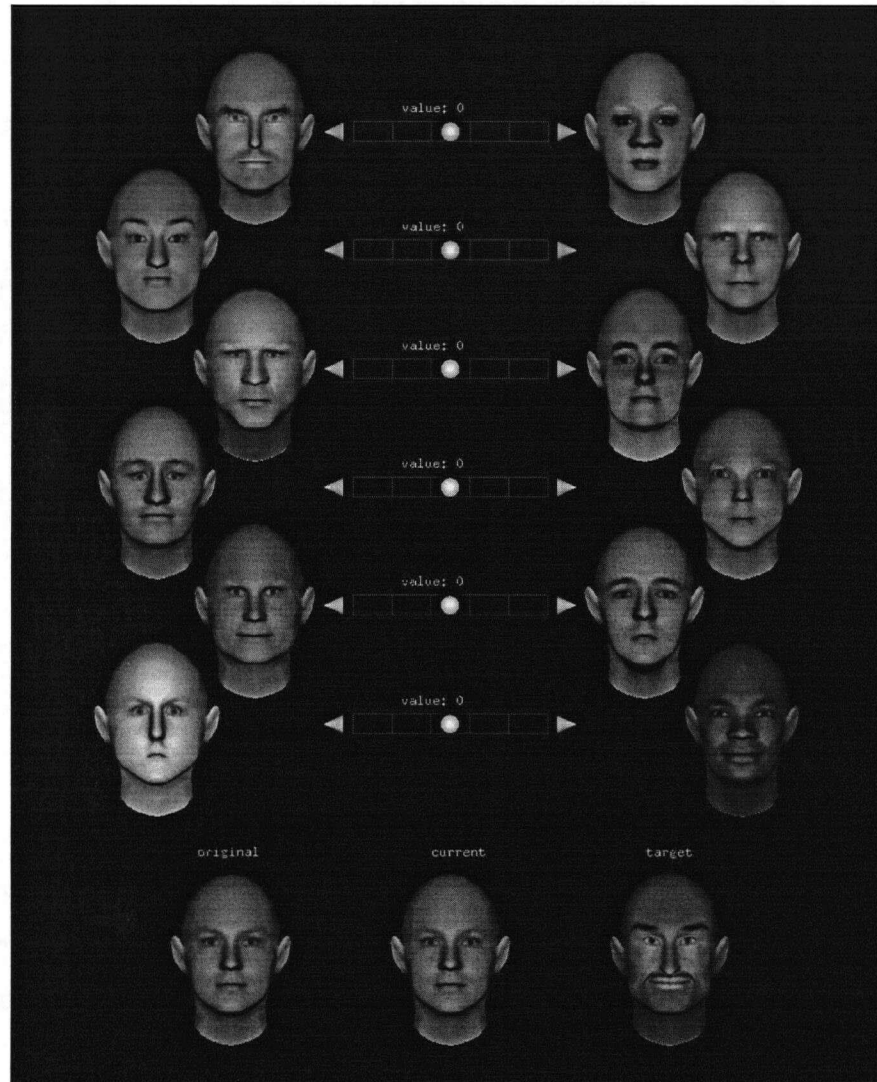
Figure 6.3: Six axes slider interface used in Pilot 1. The user is required to find the target face (shown in the lower right corner) by adjusting the sliders. Every slider has a value label above indicating the current value of the sliders, in this case, -2 to 2. The current face corresponding to the value of the sliders is shown next to the target for easy comparison. The original (average) face is shown in the lower left corner to indicate the face at the starting point.

15,625 faces. It allows users to navigate two steps in either direction along each of the six axes from the origin. We handpick six symmetric shape principal components and six symmetric texture principal components and pair them to provide a wide variety of faces. As discussed in Section 6.2.2, we adjust the axes' scales to maintain a realistic and varied face space.

For comparison, we tried to give both the wheel (Figure 6.2) and the slider interface (Figure 6.3) equal advantage. The slider interface does not show a gradient of faces like the wheel interface but allows users to slide to any desired discrete position. Consequently, quick zooming to a location and undoing a step is easy. The wheel interface, on the other hand, lacks these conveniences. The zoom-in mechanism is excluded because the database is not large enough to render it useful. However, the "back" mechanism is included so that users can undo steps if needed (a summary of the difference between interfaces is listed in Table 4.1). Although FaceGen is fast enough to create a single face during run time for our application (about 0.15s-0.2s in an Intel Pentium 4 Xeon PC), we precompute 5^6 faces for user-testing so that the wheel interface does not have the disadvantage of a time lag when displaying many faces per frame.

Each subject is provided examples of all four possible testing conditions to practise. (Pilot 1 is a 2 × 2 factorial design). Tests begin after subjects report that they are comfortable and confident in using either interfaces for face matching tasks. Each subject completes two replicates of randomly ordered treatments of all four possible conditions. Since we are not as interested in memory retention as much as closeness judgements, face targets are always dis-

played and are from the generated morphed face population. Thus, the targets actually exist in the face space the subjects navigate.

Subjects are given at most, five minutes per trial so that the experiment is short enough to end before fatigue sets in. (We adopt this time limit approach used by Schwarz et al. [52] in colour matching tasks). If the time limit is reached, a message pops up on the monitor to notify the subjects to skip the trial by pressing the space bar. It takes each subject approximately half an hour to complete the experiment. The face space locations they visited with time stamps are saved in their log files.

6.3 Results

In this experiment, subjects perform very well in conditions that use only three navigation axes. All subjects find the target face with little difficulty. Since most of the subjects find the three axes cases easy with either interface, there are no huge variations; therefore, we exclude it from further analysis other than to notice that differences occur as more dimensions are included.

We test a number of hypotheses with our data. We have also interviewed our subjects after they complete the face matching tasks to understand their preferences. Appendix A outlines the questions we ask. We present a summary of subjects' feedback after reporting our statistics.

Our hypotheses are:

- H_{01}: Subjects take similar amount of time using the wheel interface and

	Static Slider	**Wheel**
Six Axes	181.74s (69.61s)	148.02s (50.29s)

Table 6.4: Mean number of time taken (in seconds) across Pilot 1 interface conditions. Standard deviations are shown in brackets. Note that the time-out trials are excluded.

the slider interface.

- H_{02}: Subjects score similarly using the wheel interface and the slider interface.

- H_{03}: Subjects revisit faces similar number of times using the wheel interface and the slider interface.

- H_{04}: Subjects see a similar number of unique faces using the wheel and the slider interfaces.

We report our statistic results in the same order as above.

6.3.1 Performance in Time

Due to our enforcement of the five minute per trial limit, analyzing the subjects' performance based on time is inappropriate. Using the same approach as Schwarz et al. [52], we remove the time-out trials when evaluating timing. Table 6.4 shows the cell[1] means and standard deviations with time-out trials removed.

By discounting the time-out trials, we see that the slider interface takes

---

[1]The smallest unit in a factorial design is a "cell." It represent a specific condition in the experiment.

182s on average, and the wheel takes 148s on average. The difference is not significant, but it indicates a trend for cases where subjects do complete the task. In this case, the wheel interface is faster. However, more subjects do not complete the task using the wheel interface. This suggests that there may be individual differences in how subjects use the interface. It may be that some subjects are better at one interface than the other, and vice versa. This may be an interesting research direction as it may suggest that people have different mechanisms for analyzing faces. Further applications of an unbalanced one-way ANOVA to the data, excluding the time-out trials, show no significance regarding the independent variables. Thus, we cannot reject H_{01} for Pilot 1.

6.3.2   Performance in Score

We assign a score to each trial by measuring how many steps subjects' final positions are from the target. The testing scores show that all subjects successfully match faces for the wheel and slider trials with three navigation axes. For six navigation axes cases, on the other hand, four out of twelve trials of the wheel interface, and two out of twelve of the slider interface are not complete. However, four out of six subjects report that they are more comfortable using the wheel interface than the slider interface.

Again, due to the five minute time restriction, it is inappropriate to analyze the score data in ANOVA. The reason is that the time-out trials cannot count towards subjects' failure. Subjects do not complete the task because they do not have enough time, not because they are incapable. Despite this drawback,

we run an unbalanced one-way ANOVA test on the data excluding the time-out trials. The test does not show significance, therefore we cannot reject H_{02} for Pilot 1.

6.3.3 Navigation Patterns

Seeing that time and score are not meaningful dependent measurements due to the five minute restrictions, we look into subjects' navigation pattern. For the six axes interfaces, we notice that there is significantly less "wandering around" of subjects with the wheel interface. When plotting the navigation path history against the spatial distance from the target, the patterns for wheel observations often show a steady convergence to the target face with little zigzagging back and forth along a particular axis (for example, Figure 6.4). In contrast, looking at the navigation patterns for slider cases show many zigzags before the face target is reached (for example, Figure 6.5). We expect that this is due to subjects performing trial-and-error searches to discover the impact a particular slider has on the overall figure given their current position in the face space. This is why we look into repetitions of subjects' navigation.

6.3.4 Performance in Reducing Repetition

In terms of the number of face revisits, means and standard deviations for the six axes condition are given in Table 6.5. Our data show subjects revisit a face on average 2.8 times per trial using the wheel interface, but 14.4 times per trial using the slider interface.

subject 3, replicate 2, 6-D wheel
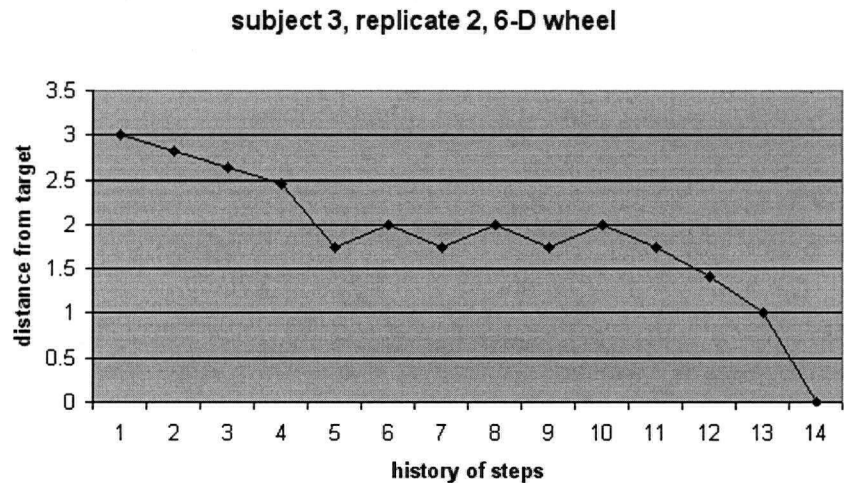
Figure 6.4: One wheel navigation pattern of subject 3 from Pilot 1. There are a few zigzag patterns.

	Static Slider	Wheel
Six Axes	14.42	2.75
	(10.11)	(5.56)

Table 6.5: Mean number of repetitions across Pilot 1 interface conditions. Standard deviations are shown in brackets.

The number of repeats shows significance in ANOVA tests (see Table B.3). With $F(1, 17)=11.29$, $p<0.01$ for the interface variable, it implies that there is a significant difference between the means of repetitions via the wheel and the slider interface. More specifically, the wheel interface has fewer repetitions. This test provides sufficient evidence to reject H_{03} for Pilot 1.

subject 3, replicate 2, 6-D slider
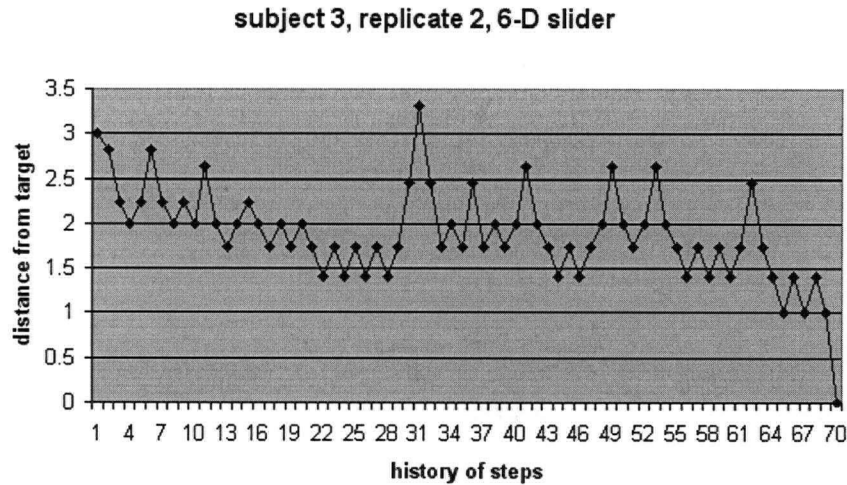
Figure 6.5: One slider navigation pattern of subject 3 from Pilot 1. Notice the frequent zigzagging within 70 moves made by the subject.

	Static Slider	**Wheel**
Six Axes	35.83	123.83
	(7.74)	(48.65)

Table 6.6: Means and standard deviations of unique faces seen across Pilot 1 interface conditions.

6.3.5 Performance in Number of Unique Faces Visited

We further analyze how many morphed faces subjects go through per trial. The means and standard deviations for the six axes condition are given in Table 6.6. We find that on average, subjects see 36 unique faces per trial with the sliders interface, compared to 124 faces per trial with the wheel.

Taking time into consideration, we also find that subjects spend an average of 5.6s per face with the slider interface, which is significantly more compared to the 1.7s per face with the wheel interface (see Table 6.7). This is not so

Subjects	Replicate	Unique Faces		Time Taken		Seconds/Face	
		Slider	Wheel	Slider	Wheel	Slider	Wheel
1	1	46	111	298.84	300*	6.49	2.70
	2	38	91	300*	148.63	7.89	1.63
2	1	32	141	171.69	300*	5.37	2.13
	2	37	133	300*	214.24	8.11	1.61
3	1	31	93	109.75	121.84	3.54	1.31
	2	49	139	234.47	192.54	4.79	1.39
4	1	41	117	116.18	74.29	2.83	0.63
	2	23	267	70.17	300*	3.05	1.12
5	1	26	94	177.33	96.58	6.82	1.03
	2	31	108	215.23	138.28	6.94	1.28
6	1	41	97	244.56	300*	5.96	3.09
	2	35	95	179.19	197.79	5.20	2.08
Total		430	1486			66.92	20.01
average		35.83	123.83			5.58	1.67

Table 6.7: Time spent by subjects on morphed faces using six navigation axes from Pilot 1. Note that * indicates time-out trials.

surprising since for any given step in the wheel interface many new faces appear on the screen at once. The expectation is that subjects can integrate knowledge of facial dimensions in the interface quickly as people are good at looking at faces and seeing differences. Thus, even when more faces are displayed, subjects should be good at filtering the important information without a significant time penalty. The wheel interface depends upon this ability for it to work as there are many faces displayed at any given time.

This variation is confirmed in ANOVA (see Tables B.4 and B.5) and both tables show the interface to be a significant factor at $F(1, 17)=38.42$, $p<0.01$ and $F(1, 17)=85.35$, $p<0.01$ respectively. These results provide sufficient evidence to reject H_{04} for Pilot 1. In other words, the wheel interface shows more unique faces for the time given when compared with the static slider interface. This, however, does not imply subjects find face targets faster using the wheel interface. We are simply using the number of unique faces and the amount of time subjects spent per trial to quantify their navigation patterns.

As a safety measure, we have further checked our data from this experiment. We do not find any significant learning effects.

To summarize what we have reported so far: the important things we learn from Pilot 1 statistics is that three navigation axes are not a suitable condition to test, as subjects find navigating them extremely easy and their performance show little variations. Also, the wheel interface surpasses the static sliders by reducing the number of revisits and provides more unique faces for selection.

6.3.6 Comments from the Interview

In general, subjects find the user testing to be a pleasant experience. They laugh when they see funny looking faces or faces they know. Two out of the six subjects prefer static sliders because it provides them with a "higher level" view. They find the wheel interface, on the other hand, more like a magnifying glass which only "augments" local regions. Four of the six subjects prefer the wheel interface because they find it gives a better sense of space. One subject mentions that he spends more time looking when using the wheel interface and spends more time clicking when using the static sliders. Another subject mentions that because the target face is at the bottom right corner, he constantly needs to look up and down to compare the face options with the face target. He feels that distance between faces can affect his comparing, and it will be useful to allow users to drag the target face around for side-by-side comparisons. All subjects find navigating with six axes to be more challenging than three axes.

6.4 Discussion

Our results show that the interface factor is only significant in the number of revisits by subjects and the number of unique faces subjects saw. (Results indicated by rejection of H_{03} and H_{04}; see Table B.2 for summary).

6.4.1 Effect of Wheel Interface

Since the wheel provides face gradients, subjects spend less time on average studying faces. They feel "comfortable" that all the immediate possible options are presented to them in the neighbourhood so that they can concentrate on those closer in similarity to the face target and ignore those that are obviously too different from the target. We speculate this to be a sign that the wheel interface is useful for face matching refinement.

6.4.2 Effect of Static Slider Interface

In contrast, the static slider interface requires users to discover the face gradients themselves. This forces them to resort to trial and error. It takes them more time to look at every face on average because they are revisiting faces they have already seen from trial and error, not because they are examining faces carefully every time they make a move. The significance of this effect is shown in the ANOVA results in Tables B.4 and B.5 of Appendix B where the interface variable is significant.

6.4.3 Navigation Patterns

Although our data do not show a significant difference in terms of the time and score of the subjects' performance, it does show two very different navigation patterns. Subjects use different strategies with the two interfaces and we believe that given more navigation axes, subjects will have a more difficult time reaching face targets by trial-and-error via sliders. In contrast, the wheel interface is able to present users with more combinations of faces for examination and to reduce users floundering in the face space. In light of this, it may be a good idea to use multiple wheels to manage the screen space effectively.

Why is it then that the wheel interface is not always faster? We suspect that subtle facial differences are hard to detect when all neighbouring faces are dominated by one very distinct look (for example, Figure 6.6). It takes more time than what is available for subjects to become familiar with these faces. This may be due to the "other race effect" [60] and depends upon the limitations of people's face recognition ability [35].

Some subjects also comment that the wheel interface is similar to a magnifying glass, forcing users to look at local areas, which is something new. However, providing the zooming facilities we develop would help this latter problem. Given a majority of the subjects prefer the wheel interface over static sliders, its usefulness is supported.
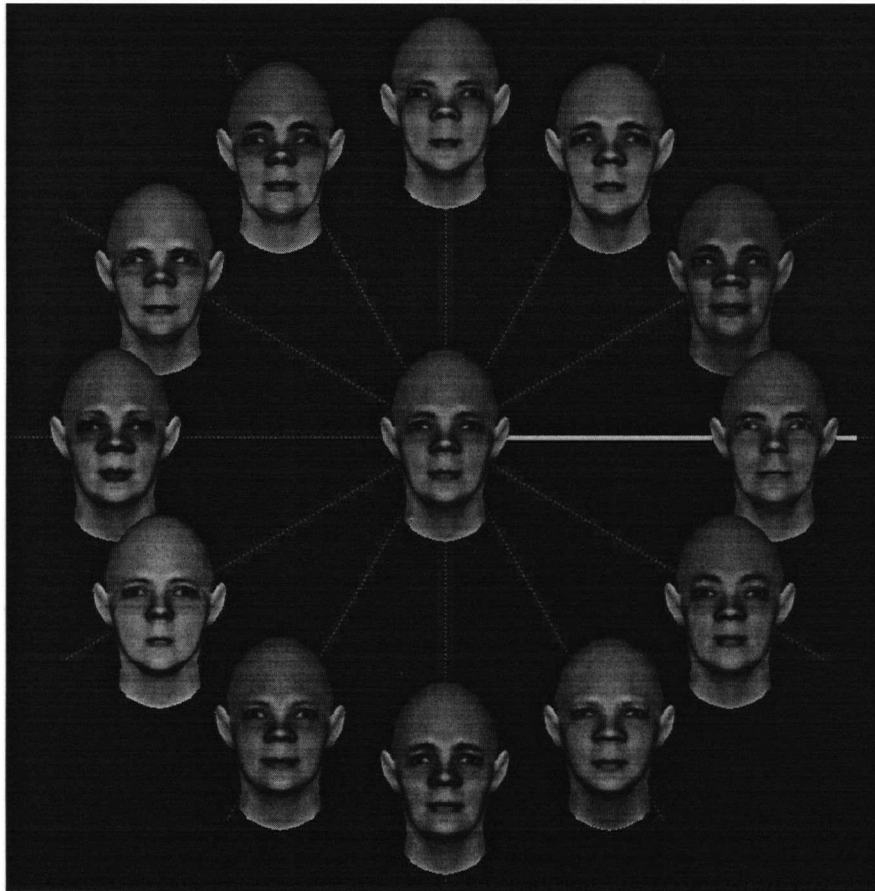
Figure 6.6: Face spectrum from Pilot 1 with similar looking faces.

6.4.4 Reminiscent of Sanjusangen-do Temple

An interesting observation of subjects' experience worth noting is that the face targets tend to look unreal and unattainable to them at first, but as navigators approach the neighbourhood of the targets, the targets look increasingly real and attainable. Perhaps it is the familiarization with faces that changes people's perception. First, the faces look like strangers; after a while, they look like friends. This echoes the feel of the Sanjusangen-do Temple. In the beginning,

all the Buddha statues look the same, but after a while they all look different.

We have given an account of the first pilot test in this chapter. We outline our method, report the results and discuss the findings. In the next chapter, we present the second pilot test we conduct.

Chapter 7

Pilot Test 2: Effect of Interface and Axes Type

7.1   Introduction

In Pilot 2 we want to investigate two issues, the performance of dynamic sliders versus our wheel interface and the difference between using axes that vary along independent principal components and correlated principal components. We use a two-variable within-subject design (Table 7.1).

We do not expect a large difference between the wheel interface and the hybrid dynamic slider interface. The reason is that they both update faces dynamically with respect to the current face space location. Hence we will not analyze their navigation pattern. We also do not expect a large difference between correlated and independent navigation axes. This is because, like colours, Schwarz has not found a large difference between people's preference with various colour representations when engaged in colour matching tasks[52].

<table>
<tr><td></td><td colspan="3" align="center">*Navigation Axes Type*</td></tr>
<tr><td></td><td></td><td>independent</td><td>correlated</td></tr>
<tr><td rowspan="2">*Interface*</td><td>**Wheel**</td><td></td><td></td></tr>
<tr><td>**Dynamic Slider**</td><td></td><td></td></tr>
</table>

Table 7.1: Two variable factorial design for Pilot 2.

7.2 Method

7.2.1 Subject

We used a new group of six graduate student volunteers from Computer Science and Electrical Engineering. Of the six, one is a woman, and five are men; and two are Caucasian, one is East Indian, and three are Oriental.

7.2.2 Apparatus

Details of the wheel and the dynamic slider interfaces are given in chapter 4. Customization for this experiment are the new control panel, target selection and navigation axes.

Control Panel

Figure 7.1 shows the control panel used for Pilot 2. The control panel works very similarly to the one used for Pilot 1 (see Section 6.2.2), except the radio buttons now indicate different testing conditions. Subjects can select any of the conditions during practise mode. If they select the "Start Test" button, the condition radio buttons are de-activated.

Face Target

Targets are chosen for both correlated and uncorrelated axes so that they require two steps along either direction of two of the axes, one step along three of the other axes and no step for one of the axes. In other words, targets are obtained

Figure 7.1: Control panel for Pilot 2.

from a random sequence of the set $\{0, 1, 1, 1, 2, 2\}$, where each element can be either positive or negative.

Navigation Axes

In Pilot 2 we compare dynamic sliders with the wheel interface under conditions of independent and correlated axes[1].

By using our full system, we are able to hand pick three symmetrical texture principal components and three symmetrical shape principal components as navigation axes and determine their realistic ranges. Table 7.2 shows three shape and three texture principal components used as independent navigation axes. Using the second row as an example, moving one step along axis A_1 is equivalent to moving a step size of five along the third symmetric shape principal component S_3—which is unlike the coupled axes used in Pilot 1 where navigating along an axis is equivalent to the manipulation of two principal components (Table 6.3). The range, and hence, step size of the navigation axes are also

---

[1]Again, we like to remind readers, as we have in Section 5.1.2, that when we refer to "independent" axes, we really mean the principal components that form the axes which are manipulated separately. Hence, they are uncorrelated in the sense of navigation axes, not in a statistical or perceptual sense.

Axes	PC	Range	Step Size
A_1	S_3	$[-10; 10]$	5
A_2	S_5	$[-10; 10]$	5
A_3	S_7	$[-10; 10]$	5
A_4	T_6	$[-8; 8]$	4
A_5	T_7	$[-10; 10]$	5
A_6	T_8	$[-6; 6]$	3

Table 7.2: Uncorrelated navigation axes used in Pilot 2.

different from the Pilot 1 axes because different principal components are chosen for this experiment and they capture different amounts of variations. Similar to Pilot 1, however, we also discretize the number of possible positions on the navigation axes to five, making the face space a total of 5^6 for each type of the navigation axes. We precompute those faces as well to speed up their update process.

For correlated axes (Table 7.3), we essentially rotate the navigation axes by 45°. Take the first row of Table 7.3 for example, navigating along axis A_1 is equivalent to navigating $\frac{5}{\sqrt{2}}$ step along S_3 in the positive direction, and $\frac{4}{\sqrt{2}}$ step along T_6 in the positive direction. We multiplied the constant $\frac{1}{\sqrt{2}}$ with the step size of each of the paired principal components so that in our predefined user face space, the distance from the origin in our defined space is kept constant whether the users navigate in correlated or uncorrelated subspace.

Note that we have used red lines as visual cues for texture axes and blue for shape axes in the uncorrelated interface (Figure 7.3 and 7.2). For a correlated interface (Figure 7.4 and 7.5), we use cyan, magenta and yellow to indicate the pairing of axes. In the slider interface, which we have introduced in the previous

Axes	PC Relation
A_1	$A_{1s} = m_1 \Delta_1 S_3, A_{1t} = n_1 \Delta_1 T_6$
A_2	$A_{2s} = -m_1 \Delta_1 S_3, A_{2t} = -n_1 \Delta_1 T_6$
A_3	$A_{3s} = m_2 \Delta_2 S_5, A_{3t} = n_2 \Delta_2 T_7$
A_4	$A_{4s} = -m_2 \Delta_2 S_5, A_{4t} = -n_2 \Delta_2 T_7$
A_5	$A_{5s} = m_3 \Delta_3 S_7, A_{5t} = n_3 \Delta_3 T_8$
A_6	$A_{6s} = -m_3 \Delta_3 S_7, A_{6t} = -n_3 \Delta_3 T_8$

Table 7.3: Parameterisation of correlated navigation axes used in Pilot 2. Variables m_i and n_i, where i is the index of the axis pair, represent some scalar values. Variables Δ_i represent the increment along the axes.

chapter, notice how the end faces of each pair of sliders in Figure 7.5 shows the face with its anti-face for one slider (for example, the top slider), but with the shape switched for its slider counterpart (the second slider, for example). This is the positive and negative correlation effect.
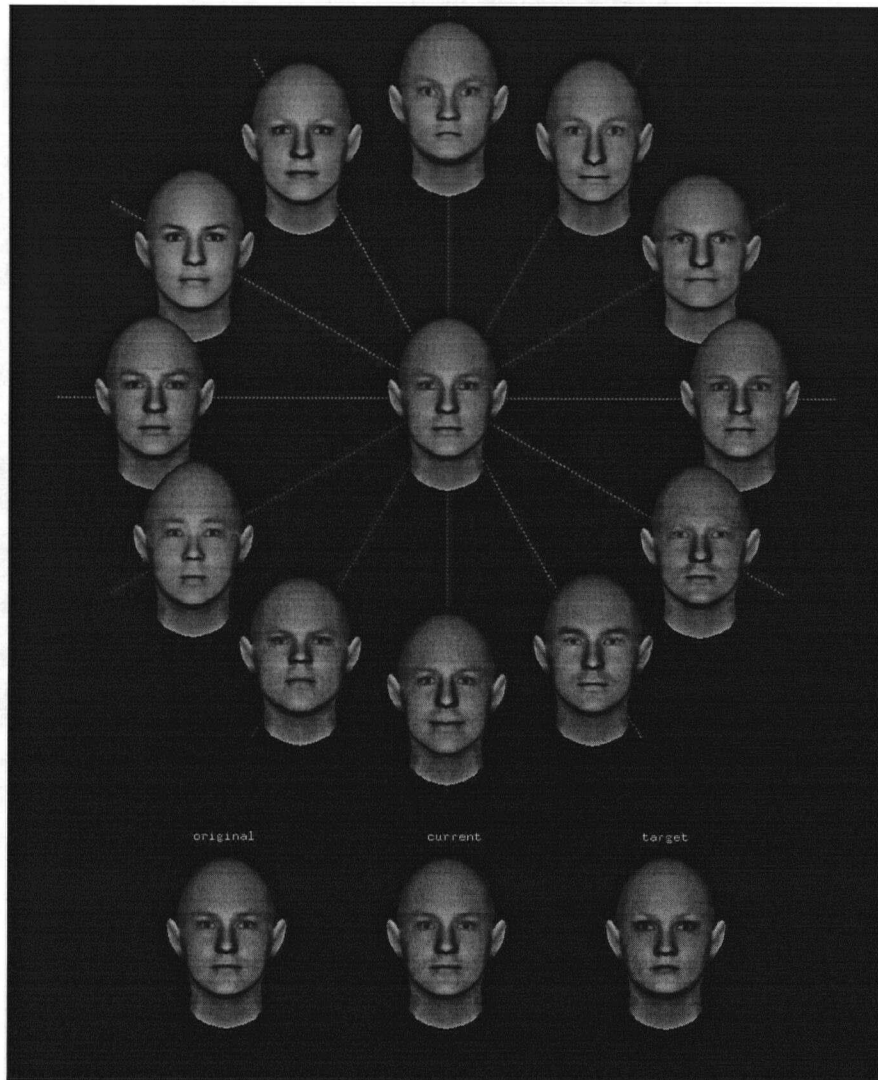
Figure 7.2: Six independent axes wheel interface used in Pilot 2. The texture axes connecting faces at 3 to 5 o' clock and 9 to 11 o'clock are outlined in red. The other three axes, outlined in blue, are shape axes.

Figure 7.3: Six independent axes slider interface used in Pilot 2. The top three sliders are outlined in red to indicate they represent texture axes. The bottom three sliders are outlined in blue to indicate they represent shape axes.
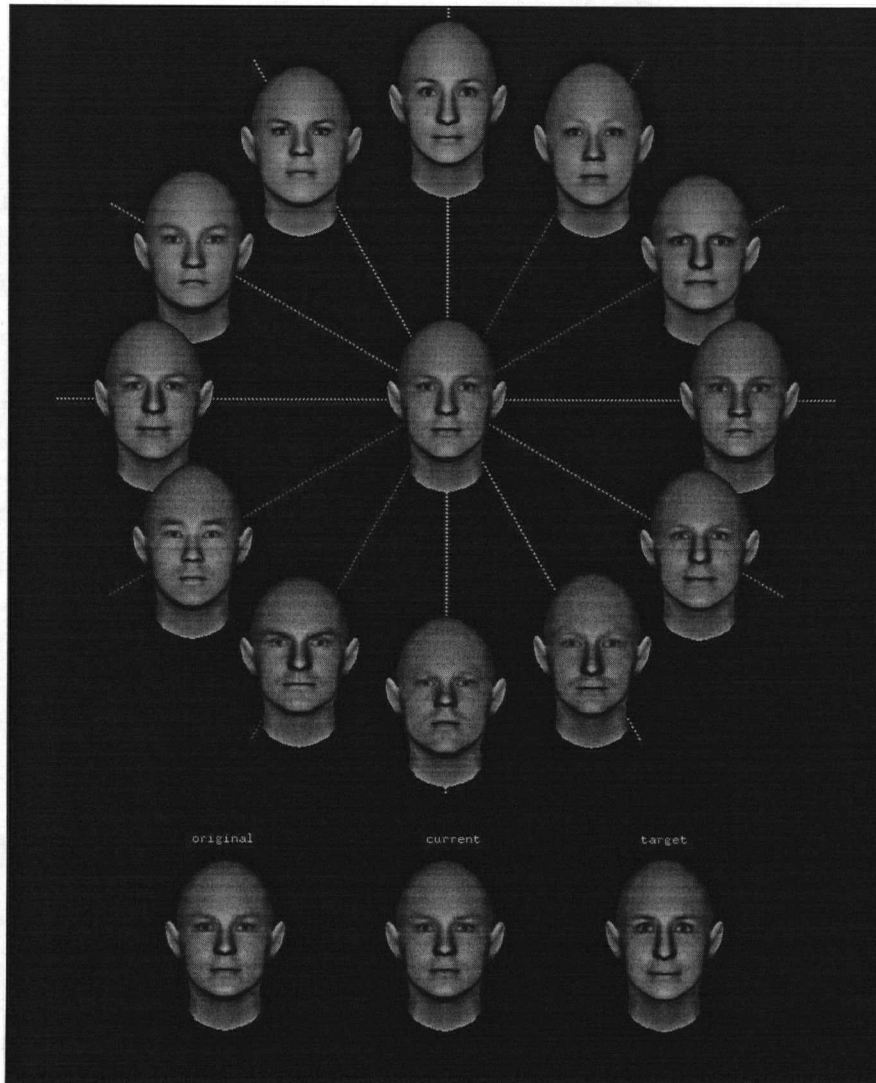
Figure 7.4: Six correlated axes wheel interface used in Pilot 2. Magenta, yellow and cyan lines are used to indicate the pairing of axes: magenta axes connect faces at 1, 2, 7 and 8 o'clock, yellow axes connect faces at 5, 6, 11 and 12 o'clock, and cyan connect faces at 3, 4, 9 and 10 o'clock.

Figure 7.5: Six correlated axes slider interface used in Pilot 2. The top two sliders have cyan outline, the middle two have yellow outlines and the bottom two have magenta outlines.

7.2.3 Procedure

In this experiment, we fix the number of navigation axes at six because we see from Pilot 1 that three axes are insufficient in showing variations of subjects' performance. We used different navigation axes as those in Pilot 1 to investigate the effect of navigation axis type. Similar to Pilot 1, we pre-generate faces with each type of axes. They populate the correlated and uncorrelated face subspace, making a total of 31,250.

We use dynamic sliders instead of static sliders for comparison. Knowing that static sliders were not as efficient as the wheel interface for subjects to search faces, we want to further narrow the gap between them by using the hybrid dynamic sliders.

In contrast to Pilot 1, the five minute limit is not strictly enforced. That is, subjects are notified upon exceeding the five minute time limit, but they can continue if they want to or move onto the next trial. Thus, subjects can take as long as they need.

Subjects are all provided with ample time to practise before testing, and each take approximately half an hour. The face locations they visited and the time stamps are recorded in their log files.

7.3 Results

We test for the following hypotheses:

- H_{01}: Subjects take a similar amount of time using the wheel interface and

	Uncorrelated Axes	Correlated Axes
Wheel	184.69s (119.17s)	213.30s (157.76s)
Dynamic Slider	175.94s (112.91s)	210.51s (142.24s)

Table 7.4: Mean and standard deviation of time taken per trial in Pilot 2.

the slider interface.

- H_{02}: Subjects score similarly using the wheel interface and the slider interface.

- H_{03}: Subjects revisit faces a similar number of times using the wheel interface and the slider interface.

- H_{05}: Subjects visit a similar number of faces during the "refinement phase" of face matching tasks, using both the wheel interface and the slider interface.

We present the statistic results in the same order of the hypotheses. Following that, comments from subjects is summarized in the interview section.

7.3.1 Performance in Time

Our data show that for an uncorrelated wheel interface, subjects require 184.7s on average. For uncorrelated sliders, however, subjects require less time, that is, 175.9s on average. As for the correlated wheel interface, subjects require 213.3s on average, and 210.5s on average with the correlated slider interface. Table 7.4 presents the cell means and standard deviations.

	Uncorrelated Axes	Correlated Axes
Wheel	100% (0%)	98% (8%)
Dynamic Slider	98% (6%)	92% (15%)

Table 7.5: Mean and standard deviation of scores in Pilot 2.

An analysis of variance on the time data does not show significant variations. This means H_{01} for Pilot 2 is not rejected.

7.3.2 Performance in Score

In terms of matching success, two of the uncorrelated slider trials are incomplete, while all trials with an uncorrelated wheel are completed successfully. As for the correlated wheel, there is one incomplete trial, but there are three incomplete trials with the correlated slider interface. Table 7.5 shows the cell means and standard deviations.

ANOVA test shows no significance with score as the dependent measurement. Thus, there is not sufficient evidence to reject H_{02} for Pilot 2.

7.3.3 Performance in Reducing Repetition

We do not notice a large difference in subjects' navigation patterns using the wheel and the dynamic slider interface. However, in analyzing the number of "revisits" subjects have (Table 7.6), we notice that on average, subjects revisit certain locations in the face space 6.6 times when they use the uncorrelated wheel interface. They revisit faces 5.3 times on average with uncorrelated sliders.

Subject	Replicate	Uncorrelated Wheel	Uncorrelated Sliders	Correlated Wheel	Correlated Sliders
1	1	0	1	0	14
	2	4	0	2	103
2	1	3	27	0	28
	2	4	7	0	1
3	1	0	0	0	18
	2	6	0	1	1
4	1	0	4	0	11
	2	0	0	2	2
5	1	38	0	33	20
	2	6	7	14	8
6	1	12	16	13	0
	2	6	1	4	0
Total		79	63	69	206
Average		6.58	5.25	5.75	17.17
Standard Deviation		10.51	8.37	9.91	28.55

Table 7.6: Number of repetitive visits by subjects over four possible conditions in Pilot 2.

	Uncorrelated Axes	Correlated Axes
Wheel	7.5 (11.35)	6.7 (10.68)
Dynamic Slider	6.2 (8.92)	8.9 (10.04)

Table 7.7: Mean and standard deviation of the number of repetitions in Pilot 2 with Subject 1 excluded.

They revisit 5.8 times on average using the correlated wheel and 17.2 times on average with the correlated sliders. Figure 7.6 shows the histogram of the average number of face revisits.

Notice in Table 7.6, the second replicate of Subject 1 under correlated slider shows 103 repetitions. This is clearly an outlier as it makes up half of the total. We exclude Subject 1 from the data analysis and the average number of face revisits taken by subjects now look like Figure 7.7. The new cell means and standard deviations are shown in Table 7.7.

Average number of face revisits
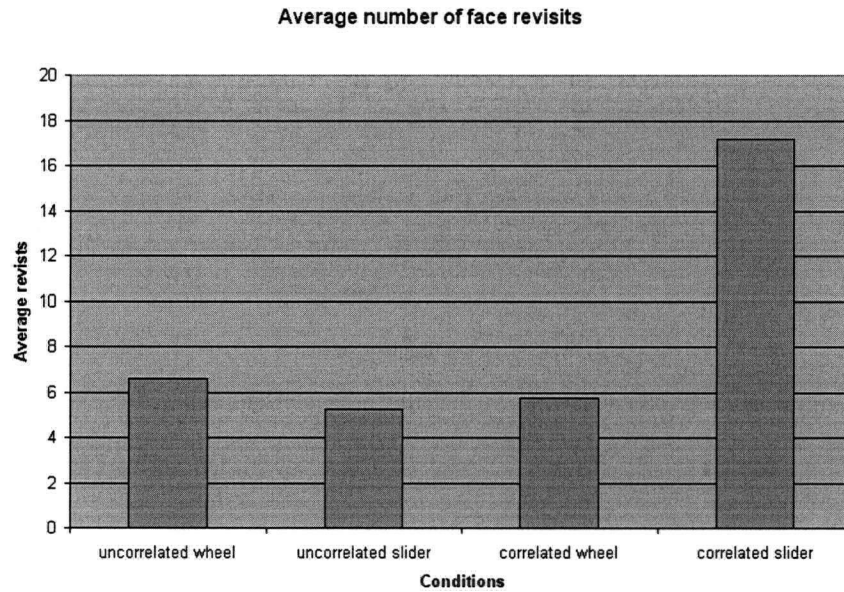
Figure 7.6: In Pilot 2 subjects revisit faces most frequently under the condition of using dynamic sliders with correlated axes.

The histogram in Figure 7.7 still shows correlated sliders to have the highest amount of repetition although the difference is not as pronounced. This is an indication that perhaps dynamic sliders are not a particularly effective interface for navigating with correlated axes. Given that static sliders are not as powerful as dynamic sliders, we anticipate that subjects would not do very well using static sliders with correlated axes either.

ANOVA tests with the repetition data with and without Subject 1 show no significance of any variables; hence interface is not a source of substantial variation, and we have insufficient evidence to reject H_{03} for Pilot 2.

Average number of face revisits - without subject 1
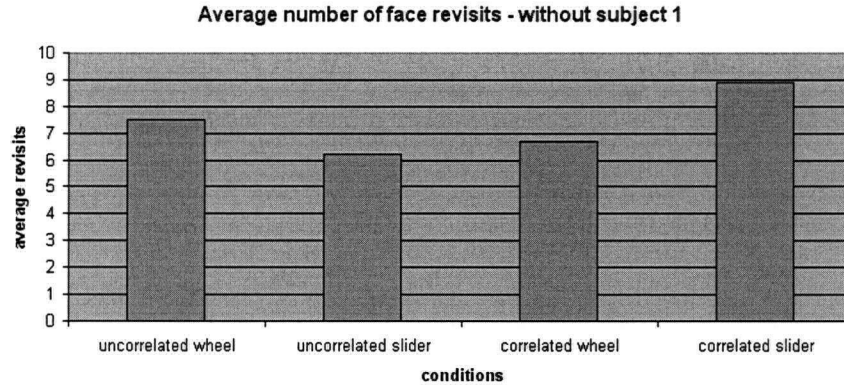
Figure 7.7: With Subject 1 removed, the average number of revisits by subjects is still the highest with correlated sliders.

7.3.4 Performance in Refinement Stage of Navigation

Last, we study the navigation phases of our subjects. Since dynamic sliders are very similar to the wheel interface in that they both show a gradient of faces[2], we do not see a dramatic difference in the trend of their navigation patterns. However, in summing up the number of times subjects have come within a certain distance of the target, we notice that subjects navigate a significant number of faces in the range of $[0, 2]$ steps from the target.

Table 7.8 and 7.9 show the cell means and standard deviations for refinement phase with and without Subject 1 respectively. Figure 7.8 shows the pattern with Subject 1 included. Figure 7.9 shows the pattern excluding Subject 1. We call these visits "arrivals," and we believe this is the stage during which subjects are polishing their face matches after finding the neighbourhood.

---

[2]As we mentioned in a previous chapter (Table 4.1 of Chapter 4), the wheel interface displays the immediate neighbours with respect to the current face and the dynamic slider interface displays the extreme neighbours relative to the current face.

	Uncorrelated Axes	Correlated Axes
Wheel	10.08	12.91
	(6.40)	(12.87)
Dynamic Slider	12.50	28.42
	(13.81)	(34.60)

Table 7.8: Mean and standard deviation of the number of faces in refinement region in Pilot 2 including subject 1.

	Uncorrelated Axes	Correlated Axes
Wheel	10.60	14.30
	(6.75)	(13.77)
Dynamic Slider	14.20	19.50s
	(14.62)	(17.58)

Table 7.9: Mean and standard deviation of the number of faces in refinement region from Pilot 2 with Subject 1 taken out.

What is interesting about these graphs is that the correlated slider interface causes the largest refinement—as indicated by the red dash lines. We further seek for evidence by applying ANOVA to the data with and without data from Subject 1. Although ANOVA results do not show any significant variables for the number of face locations visited in refinement region (hence we cannot reject H_{05} for Pilot 2), Figures 7.8 and 7.9 are an indication that subjects are having more trouble reaching the targets when using the correlated slider interface, even when they are in the neighbourhood. This implies that the wheel interface is probably better for refinement when the navigation axes are correlated.
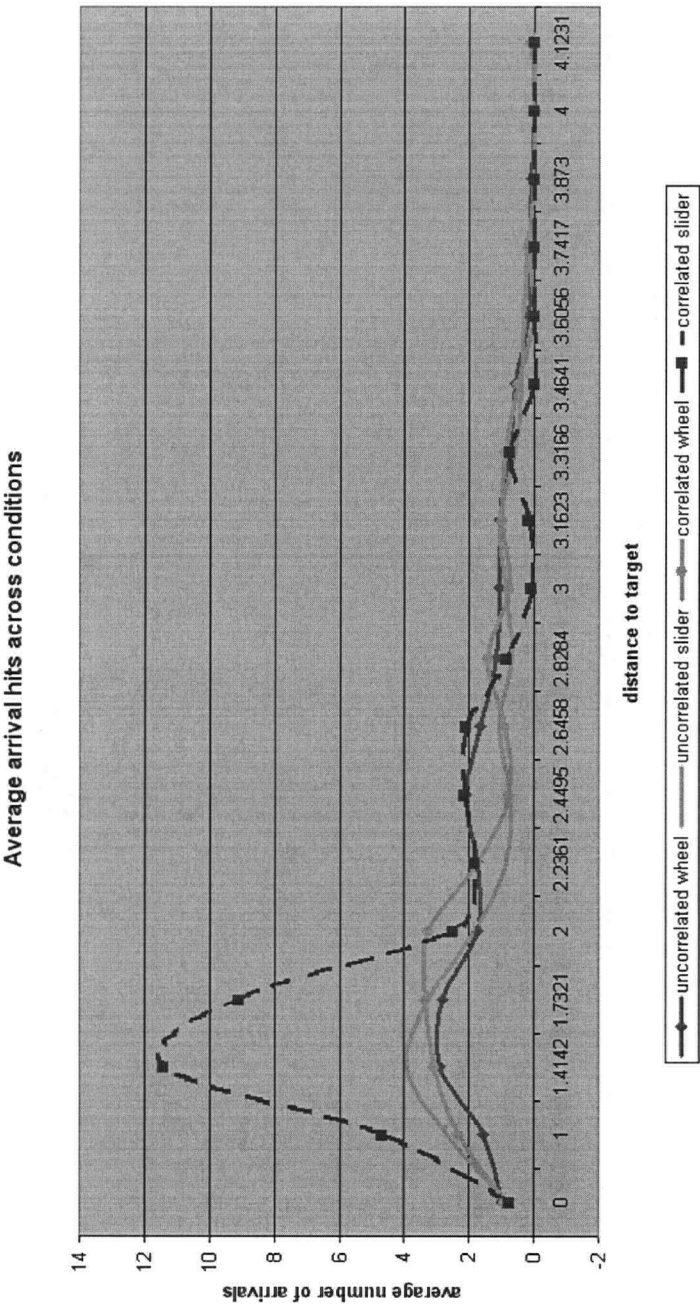
Figure 7.8: Including Subject 1, the red dash line in the graph above indicates that subjects, in refining their face matches in Pilot 2, have greater "wandering" effect when using the correlated sliders. They are within two steps of the target, almost three times as often as in the other experimental conditions, but do not seem to be aware that they are so close to the target.
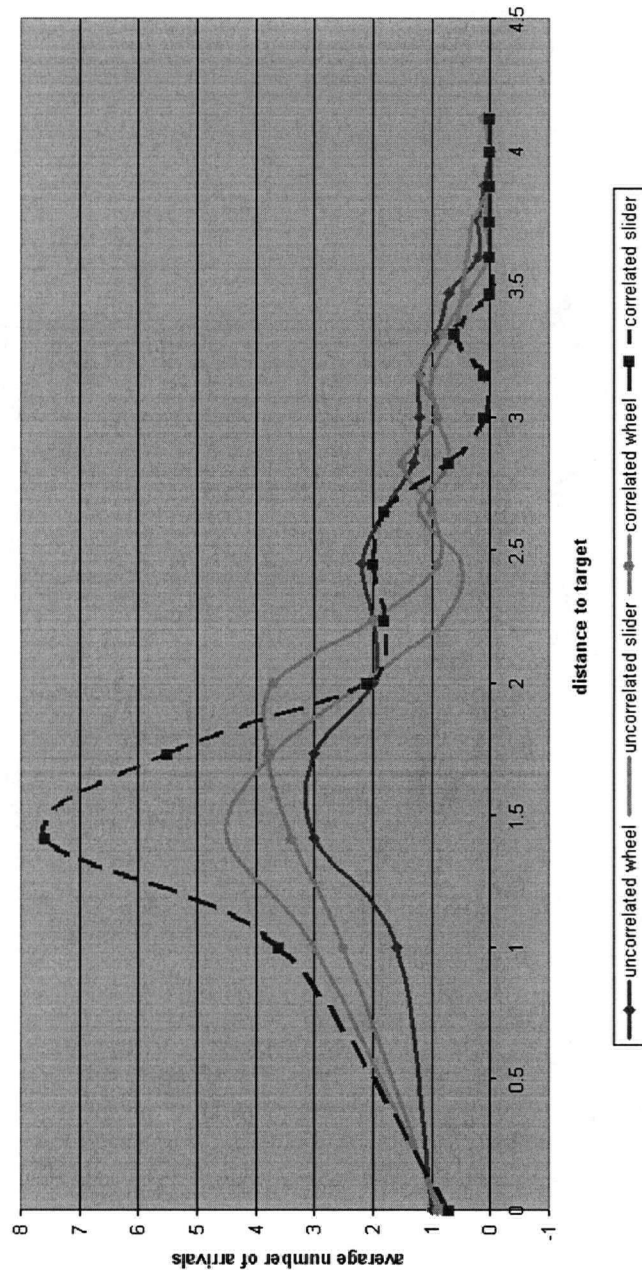
Figure 7.9: Excluding Subject 1, the red dash line in the graph above indicates that subjects, in refining their face matches in Pilot 2, still have greater "wandering" effect when using the correlated sliders. The difference is not as pronounced as in Figure 7.8 but the effect is present.

As a safety precaution, we have also compared the data of the replicates and do not detect any significant learning effects.

The noteworthy findings from what we have reported so far is that dynamic sliders and the wheel interface do not differ much in their navigation patterns. Although ANOVA tests reveal no significance of any variables no matter what the dependent measurement is, we notice that the wheel interface seems a little better for navigation with correlated axes. This may be because gradient-based interface is naturally suited for refinement.

7.3.5 Comments from the Interview

In this experiment, most subjects appear to find navigating with sliders to be more difficult. One subject remarks that he feels "so close yet so far away". Four out of the six subjects prefer the wheel interface. As one subject explains it: it is like knowing what clothes to wear when the options of combined features are presented to the users. Two out of six subjects, on the other hand, prefer the dynamic slider interface. One subject suggests that it might be useful to toggle the central face of the wheel with the target face. As for the navigation axes, most of the subjects do not sense a difference between correlated and independent axes.

7.4 Discussion

In comparing users' performance when using the wheel versus the dynamic slider interface with both correlated and uncorrelated axes, an analysis of variance

does not capture any significant variables with any data measured in time, score, repetitions, refinement or approaching efficiency. Hence, no null hypotheses are rejected (see Table B.2 for summary). This outcome is expected because we are really exploring a different way of representing faces by rotating the navigation axes. Since there is very little difference in subjects' matching colours using various colour representation [52], it makes sense that we do not find any significance with faces either. The fact that subjects do not really notice the difference between the types of navigation axes from the interview also supports this. Furthermore, the wheel interface and the dynamic interface are very similar. The only difference is that the former shows immediate neighbours, while the latter shows extreme neighbours.

7.4.1 Navigation During Refinement

We do, however, notice that subjects have more trouble reaching the targets when using correlated dynamic sliders than using the others. From our observation of Figure 7.8, we notice that subjects tend to visit a greater number of faces that are within two units of the target face when using correlated sliders. Even after removal of the outlier (Subject 1), the trend still persists (see Figure 7.9), as indicated by the red dash line which outlines a distinctly higher hump in the graph.

7.4.2   Cancelling Effect

Why then do subjects have less success when navigating via the correlated dynamic sliders? This difficulty is perhaps due to the subjects' unawareness of the cancelling effect with correlated sliders. For example, moving one step along the correlated axis A_1, and moving one step along the correlated axis A_2, may result in the doubling of the texture value $(n_1 \Delta_1 T_6 + n_1 \Delta_1 T_6)$, but also in the cancelling of the shape values $(m_1 \Delta_1 S_3 - m_1 \Delta_1 S_3)$. (Table 7.3 shows the parametrization of Pilot 2 navigation axes). Figures 7.10 and 7.11 illustrate the cancelling of shape and texture respectively. In contrast, when using the wheel interface, users move one step at a time; thus they are not easily tricked by the cancelling effect. This outcome may very well imply the possibility that our perception of faces is additive rather than subtractive.

7.4.3   Effect of Navigation Axes

Note that in Pilot 1 the six navigation wheel interface (using coupled navigation axes) only costs an average of 2.8 revisits (as indicated by Table 6.5), which is roughly half of the number of revisits we receive from the wheel interface in Pilot 2 (without Subject 1, it takes 7.5 times on average for an uncorrelated wheel interface and 6.7 times on average for the correlated wheel). This difference is likely due to the navigation axes. We suspect that the coupled axes, being positively correlated, produce more prominent faces when the texture and shape principal components are paired. More explicitly, in Pilot 1, there are a total of twelve principal components which form the six navigation axes, while there are
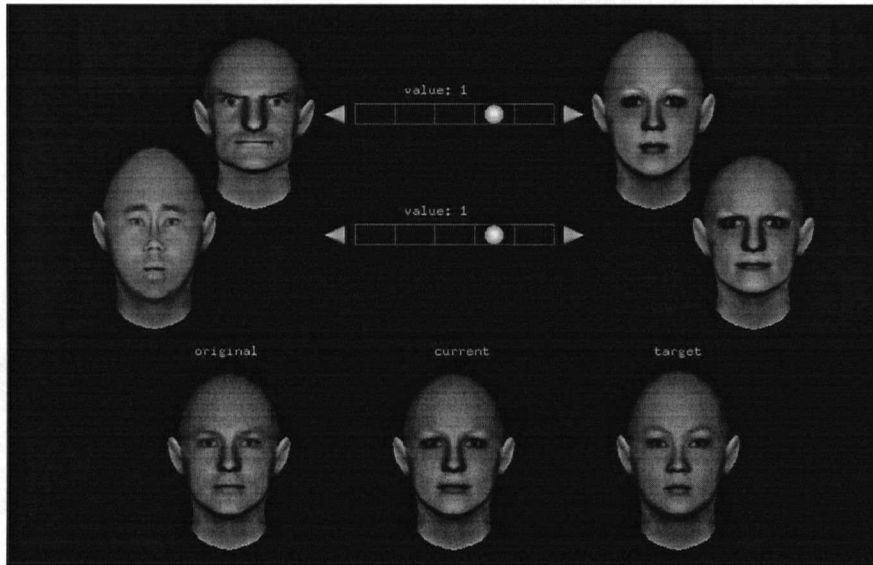
Figure 7.10: Cancelling effect results in the current face (bottom centre) with no effect on the shape and doubling of texture. In this case, only the bottom two sliders have value 1. The rest have slider value 0.

only a total of six principal components which form the two kinds of navigation axes in Pilot 2 (see Tables 6.3 and 7.2). Faces in Pilot 1 thus have more facial variation than faces in Pilot 2. This results in Pilot 1 faces being more distinct, and therefore, easier to spot. This is possibly why there are less repetitions for the wheel interface in Pilot 1.

We have described our second pilot test in this chapter. The experiment is similar to Pilot 1 but different results are obtained. We present the third pilot test next.
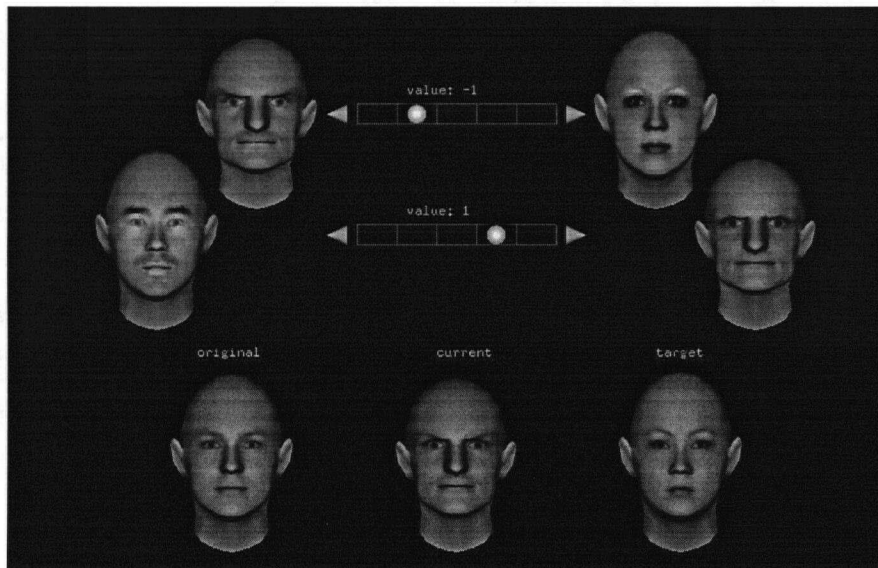
Figure 7.11: Cancelling effect can also result in the doubling of shape, but also the cancelling of textures in the current face (bottom centre). In this case, the bottom two sliders have opposite values.

Chapter 8

Pilot Test 3: Effect of Interface, Resolution and Target Distance

8.1 Introduction

In Pilot 3, we test three independent variables: interface (wheel and dynamic sliders), resolution (coarse and fine) and target type (near and far). This gives us four conditions with dynamic sliders and four conditions with the wheel interface (Table 8.1); that is, two treatments of three independent variables, giving us a total of eight different conditions.

We expect there will not be much difference between the wheel and the dynamic slider interface. However, we do expect resolution to have an effect as the finer the resolution, the harder it is to distinguish faces. We also expect target distance to have an effect. We believe the further the target, the harder it is to get there.

Distance of target from the origin

		Near	Far
Resolution	**Coarse**		
	Fine		

Table 8.1: Four experimental conditions in Pilot 3 for both the wheel and the slider interfaces.

8.2 Method

8.2.1 Subject

We recruited six graduate students from Engineering and Computer Science as subjects. Of the six volunteers, one is a woman and five are men; three are Caucasian and three are Oriental.

8.2.2 Apparatus

Customization of interfaces for this experiment are the new control panel, target selection and navigation axes. For details of the interface implementation, please refer to Chapter 4.

Control Panel

For Pilot 3, we provide a slightly different control panel compared to Pilot1 and 2 (see Figures 6.1 and 7.1). Figure 8.1 illustrates its appearance. First, it has an addition of a progress metre. Since we do not have the luxury of a gigantic amount of disk space to precompute 9^6 number of faces, we generate faces on the spot. A progress gauge notifies users of the status of their face creation, so they wait for faces to update instead of clicking repetitively.

Second, we only provide four practise conditions instead of eight for this $2\times2\times2$ factorial design experiment. This is because we do not want subjects to be aware of the target variability as it can help them "guess" the answer when being told the target is either near or far.
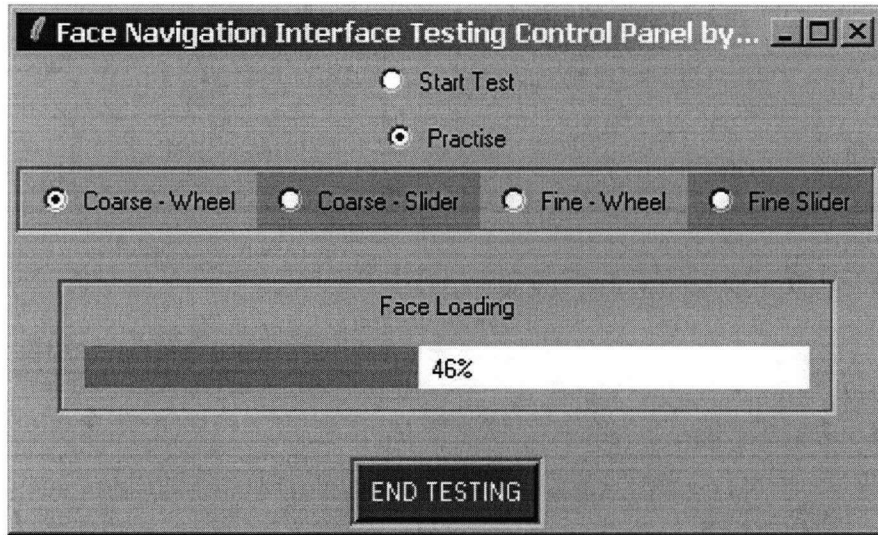
Figure 8.1: Control panel for Pilot 3 and the final experiment.

Face Target

Similar to our previous experiments, we choose our targets for each condition in such a way that subjects are required to make the same amount of moves on the navigation axes. The targets are also spatially equidistant away from the origin in the face space. Table 8.2 provides the summary.

Resolution	Distance from Origin	Set of Numbers	Example Target
Coarse	Near	$\{0,0,0,1,1,2\}$	$\{-1,0,0,1,0,-2\}$
Coarse	Far	$\{0,1,1,1,2,2\}$	$\{-1,1,2,0,-2,-1\}$
Fine	Near	$\{0,0,0,1,2,3\}$	$\{1,0,-2,0,-3,0\}$
Fine	Far	$\{0,1,1,2,3,4\}$	$\{0,-1,3,4,-2,1\}$

Table 8.2: In Pilot 3 targets are formed by the set of numbers in the third column. Each target face is any sequence of the numbers in the set which can take up a positive or a negative sign indicating directions along the axes. The fourth column provides sample targets.

Axes	PC	Range	Coarse Step Size	Fine Step Size
A_1	S_6	$[-14; 14]$	7	3.5
A_2	S_8	$[-12; 12]$	6	3
A_3	S_{10}	$[-10; 10]$	5	2.5
A_4	T_1	$[-2; 2]$	1	0.5
A_5	T_3	$[-4; 4]$	2	1
A_6	T_7	$[-10; 10]$	5	2.5

Table 8.3: Uncorrelated navigation axes used in Pilot 3 and the final experiment.

Navigation Axes

In Pilot 3, we compare the dynamic slider against the wheel interface and vary the condition of resolution and target distance. We use independent navigation axes in this case, and the details are described below.

We use independent axes of which three are symmetrical texture principal components and three are symmetrical shape principal components (similar to the uncorrelated axes condition in Pilot 2). The axes and principal component mappings are shown in Table 8.3. We list both the step size for the coarse resolution and the step size for the fine resolution. The former step size is the double of the latter so users would take twice as many steps to get to a destination when they use the fine resolution.

Snapshots of the interfaces are shown in Figures 8.2 and 8.3 for coarse resolution, and Figures 8.4 and 8.5 for fine resolution. When comparing the wheel interface for coarse (Figure 8.2) and fine (Figure 8.4) resolutions, notice that the neighbouring faces of the central face are more prominent with the coarse resolution (Figure 8.2). Similarly, juxtaposing the two slider interfaces for different resolutions, the slider interface for coarse resolution (Figure 8.3) allows

five possible values for each slider, whereas the slider interface for fine resolution

(Figure 8.5) allows nine.

We outline the testing procedure in the next section.
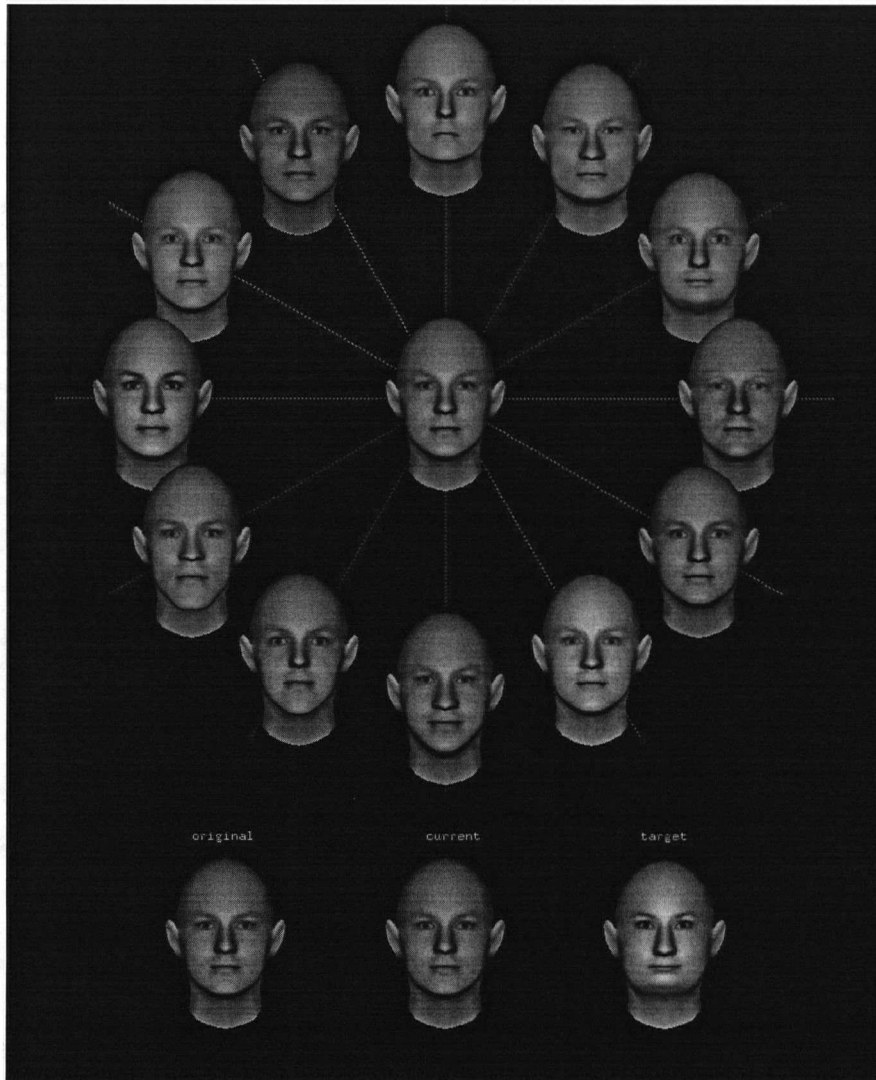
Figure 8.2: Six uncorrelated axes wheel interface used in Pilot 3 and the final experiment with coarse resolution. The shape axes, outlined in blue, connect faces at 12 to 2 and 6 to 8 'o clock. The other axes, outlined in red, are texture axes.
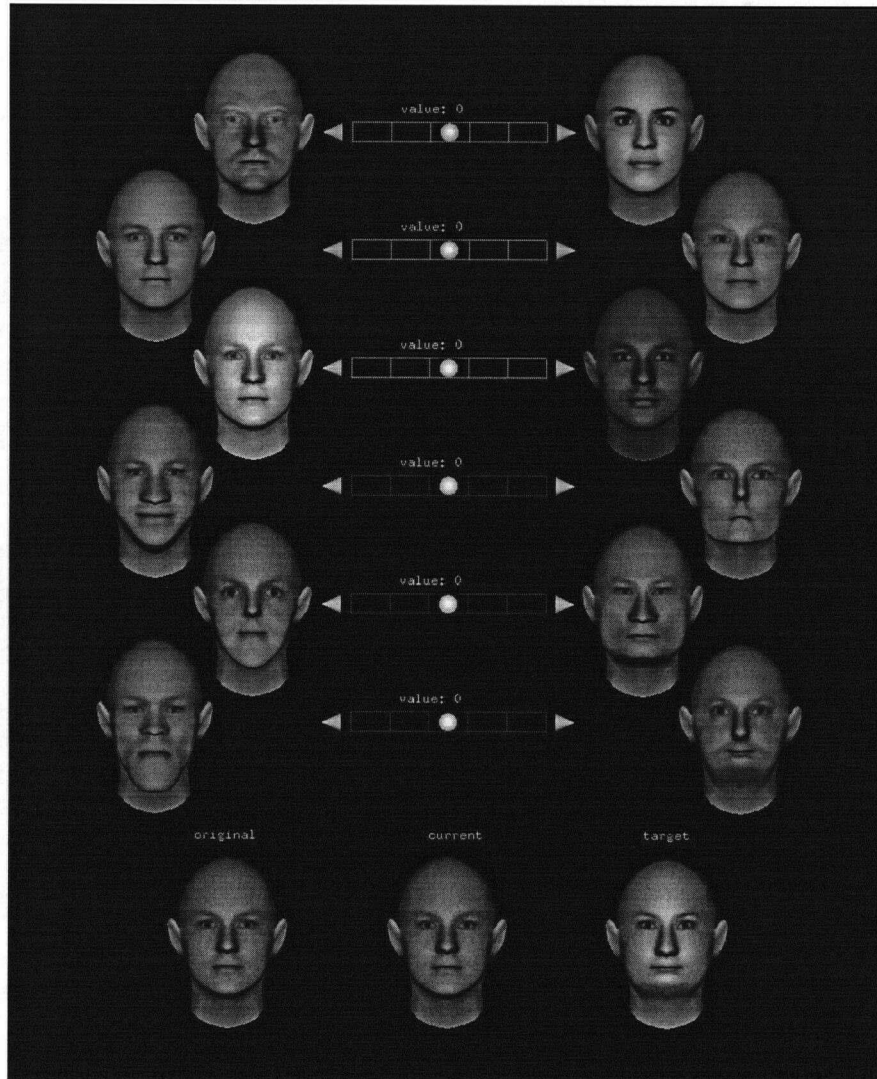
Figure 8.3: Six uncorrelated axes slider interface used in Pilot 3 and the final experiment with coarse resolution. The top three sliders are outlined in red indicating texture axes. The bottom three sliders are outlined in blue indicating shape axes.
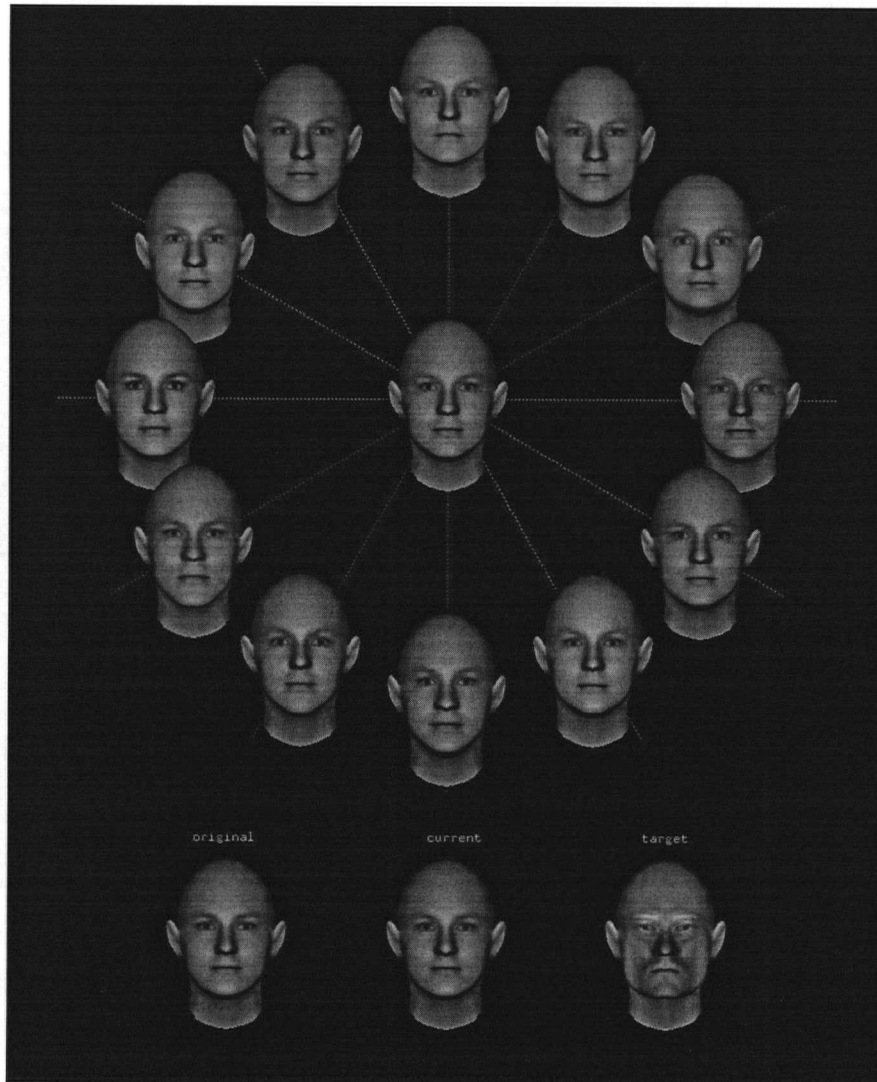
Figure 8.4: Six uncorrelated axes wheel interface used in Pilot 3 and the final experiment with fine resolution. The blue shape axes connect faces at 12 to 2 and 6 to 8 o' clock. The red texture axes connect faces at 3 to 5 and 9 to 11 o' clock.
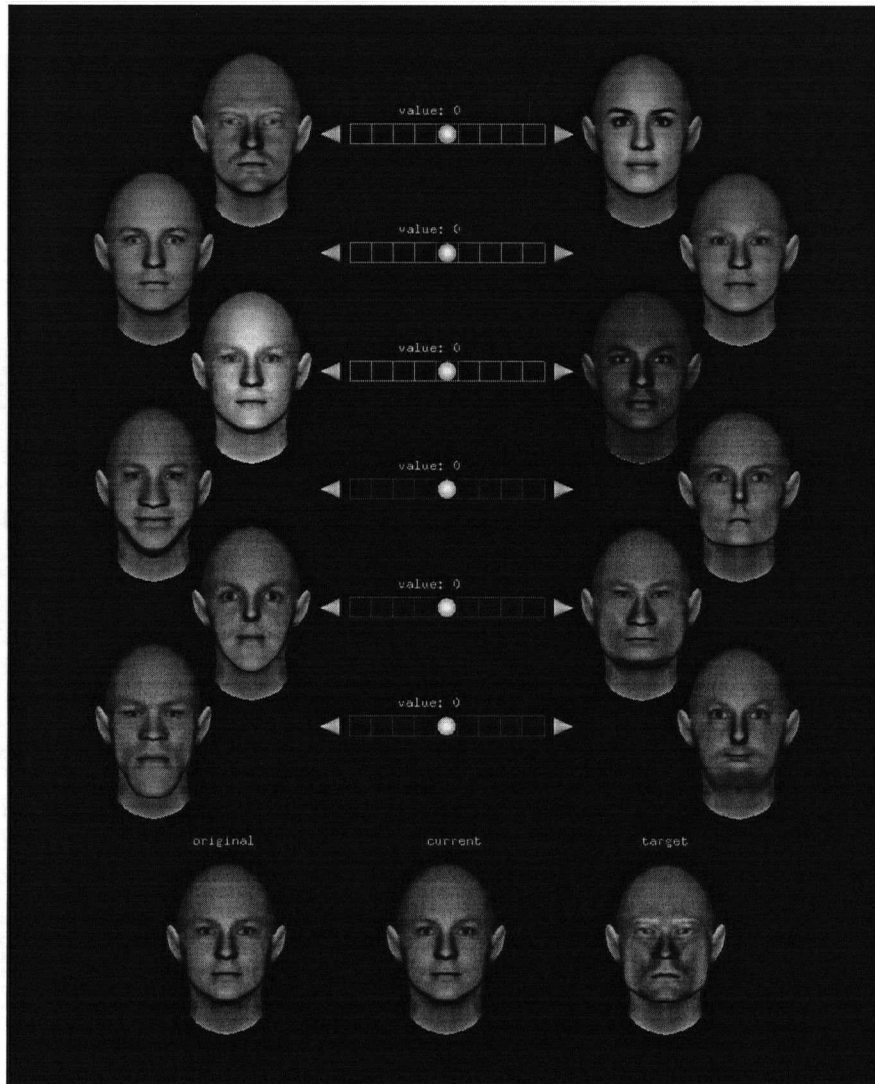
Figure 8.5: Six uncorrelated axes slider interface used in Pilot 3 and the final experiment with fine resolution. The top three red sliders indicate texture axes and the bottom three blue sliders represent shape axes. The value label above each slider displays values from -4 to 4 indicating the discrete position of each slider.

8.2.3 Procedure

In this experiment, we choose to use independent texture and shape axes as we learned from Pilot 2 that correlated axes can be potentially confusing for subjects. The previous section describes the navigation axes. We exclude the "zoom" feature of the wheel even though the face space is much bigger than before.

Unlike the previous two pilot tests, since we do not have a large hard disk to store the face space population, we generate faces at run time. If we have the disk space, the number of faces needed is 9^6 which is 531,441 in total.

Each subject has to complete sixteen trials, requiring approximately 30 minutes to one hour each to complete the test. Subjects also practise using each interface with different resolutions. However, during practise, targets are random with varying difficulties. They do not necessarily conform to the sequences listed in Table 8.2. The reason we do not allow subjects to practise with near and far targets exclusively is because we do not want them to work out the target patterns. The near targets are especially easy to figure out with ample practise. In this way, target is a "hidden" variable.

Subjects' navigation patterns are recorded by registering the locations in face space visited and the time stamps of each move they made. These are the data collected.

8.3 Results

We test the following hypotheses for this experiment:

- H_{01}: Subjects take a similar amount of time using the wheel interface and the slider interface.

- H_{02}: Subjects score similarly using the wheel interface and the slider interface.

- H_{03}: Subjects revisit faces a similar number of times using the wheel interface and the slider interface.

- H_{05}: Subjects visit a similar number of faces during the "refinement phase" of face matching tasks, using both the wheel interface and the slider interface.

- H_{06}: Subjects visit a similar number of faces during the "approaching phase" of face matching, using both the wheel interface and the slider interfaces.

Statistical results are presented in the same order as the hypotheses. However, testing of H_{05} and H_{06} are discussed together because approaching and refinement are closely related by our definition (see Section 21).

8.3.1 Performance in Time

The average time subjects require is shown in Figure 8.6. It is evident that subjects take the longest using the wheel interface under a fine resolution con-

dition when searching for a far face target (column seven of the histogram). Its counterpart (the eighth column) on the other hand, takes less time. The conditions, where subjects are looking for near face targets under a coarse condition, also shows that subjects take quite a bit of time (column three and four of the histogram). Table 8.4 shows the cell means and standard deviations.

Time taken across all conditions
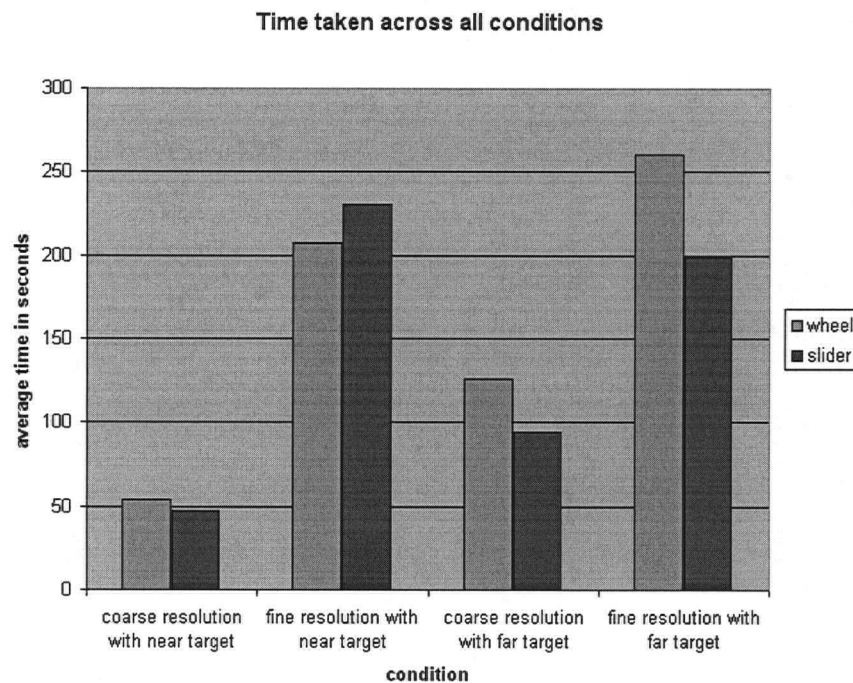
Figure 8.6: The histogram shows the average time (in seconds) taken by subjects per trial across all eight conditions of the Pilot 3 experiment.

From ANOVA table B.6, interface is not a significant factor ($F(1, 5) = 0.70$, $MS_e = 12479.60$, $p > 0.05$); thus, there is no sufficient evidence to reject H_{01} for Pilot 3. Resolution is a significant factor ($F(1, 5) = 46.28$, $MS_e = 10813.30$, $p < 0.01$), and we know the finer the resolution, the longer time subjects need

Wheel	Near Targets	Far Targets
Coarse Resolution	53.58s	126.05s
	(21.83s)	(68.45s)
Fine Resolution	207.14s	260.98s
	(125.40s)	(126.13s)
Dynamic Slider	**Near Targets**	**Far Targets**
Coarse Resolution	46.81s	94.26s
	(17.51s)	(50.73s)
Fine Resolution	230.94s	199.22s
	(151.04s)	(67.09s)

Table 8.4: Mean and standard deviation of the amount of time taken per trial in Pilot 3.

to complete the tasks (as indicated by comparing the cell means in Table 8.4). Target is not significant ($F(1, 5)$=5.24, MS_e=5780.10, $p<0.07$) and none of the two-way and three-way interactions are significant.

8.3.2 Performance in Score

In terms of the accuracy score, which we assign to each trial depending on how far away the subjects are from the target, we find that all the trials using the coarse resolution have perfect scores. More variations are detected using fine resolution. Table 8.5 shows the cell means and the standard deviations.

As shown in Figure 8.7, subjects appear to perform the worst using the wheel interface with fine resolution in search of near targets (column three of the histogram). The low score is an indication that not only do subjects fail to find face targets, they also end up somewhat further away from the targets. The counter part (using sliders) indicated in column four of Figure 8.7, in contrast, works better. When subjects use the wheel interface, with fine resolution, in search of

Wheel	Near Targets	Far Targets
Coarse Resolution	100% (0%)	100% (0%)
Fine Resolution	81% (32%)	95% (11%)
Dynamic Slider	**Near Targets**	**Far Targets**
Coarse Resolution	100% (0%)	100% (0%)
Fine Resolution	92% (19%)	96% (9%)

Table 8.5: Mean and standard deviation of the subjects' scores in Pilot 3.

a far target (column seven of the histogram), although they take longer, they come closer to the target when compared to a similar condition with near target (column three). The counterpart (column eight of the histogram) scores a little better. This is an unintuitive and unexpected outcome, as we anticipate that the closer the targets, the fewer the axes of manipulation; therefore, subjects are expected to find it easier and score higher. We discuss possible causes of this in the next section.

ANOVA test reveals no significant variables. Since interface is not significant, there is no evidence to reject H_{02} for Pilot 3.

8.3.3 Performance in Reducing Repetition

We sum up the number of revisits by subjects across all conditions to see how much repetition occurred in each condition. Table 8.6 shows the cell means and the standard deviations.

Figure 8.8 shows the distribution. This data echoes the outcome of the testing score. Subjects appear to revisit many times when required to find

Average score across conditions
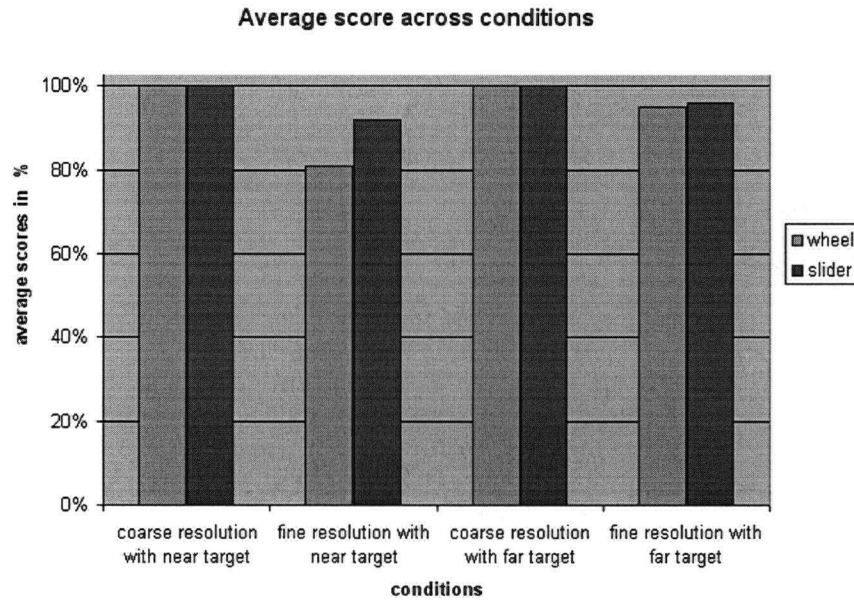
Figure 8.7: Accuracy score of subjects by trial across all eight conditions of the Pilot 3 experiment.

near targets while navigating in fine resolution (column three and four of the histogram). This indicates inefficiency as a lot of repetitive viewing of the same set of faces occurs. Subjects also exhibit the same behaviour when required to find far away targets under the fine resolution condition (columns seven and eight of the histogram). However, the number of revisits are not as high in comparison.

From ANOVA Table B.7, interface is not a significant variable ($F(1, 5)=0.01$, $MS_e=5.90$, p>0.05); therefore, there is no sufficient evidence to reject H_{03} for Pilot 3. Resolution is a significant variable ($F(1, 5)=11.40$, $MS_e=11.88$, p<0.05) affirming that the finer the resolution, the more repetition. Target is not significant ($F(1, 5)=0.01$, $MS_e=10.09$, p>0.05). Lastly, none of the interactions

Wheel	Near Targets	Far Targets
Coarse Resolution	0.25 (0.62)	0.67 (0.89)
Fine Resolution	3.67 (4.23)	2.39 (2.59)
Dynamic Slider	Near Targets	Far Targets
Coarse Resolution	0.08 (0.29)	1.33 (2.39)
Fine Resolution	3.50 (3.23)	2.25 (2.09)

Table 8.6: Mean and standard deviation of repetitions in Pilot 3.

between the variables are significant.

8.3.4 Performance in Approaching and Refinement

We further analyze the navigation pattern and see how many times subjects come within a certain distance of the target. Figures C.1, C.2, C.3 and C.4 in Appendix C show the plots of the two interfaces across four possible conditions (coarse resolution with near target, fine resolution with near target, coarse resolution with far target and fine resolution with far target). We notice that the magenta line (representing the sliders) seems to have the tendency of being below the blue line (representing the wheel) when the distance from the target is more than two. When the distance is within two steps, however, the magenta line seems to have the tendency to surpass the blue line. This trend appears some what irregular, but it deserves a closer look.

In order to understand how subjects use both interfaces, we compare the percentage of face locations that subjects visited that fall within the refinement category with the percentage of face locations visited that fall within approach-

Average number of repeats across all conditions
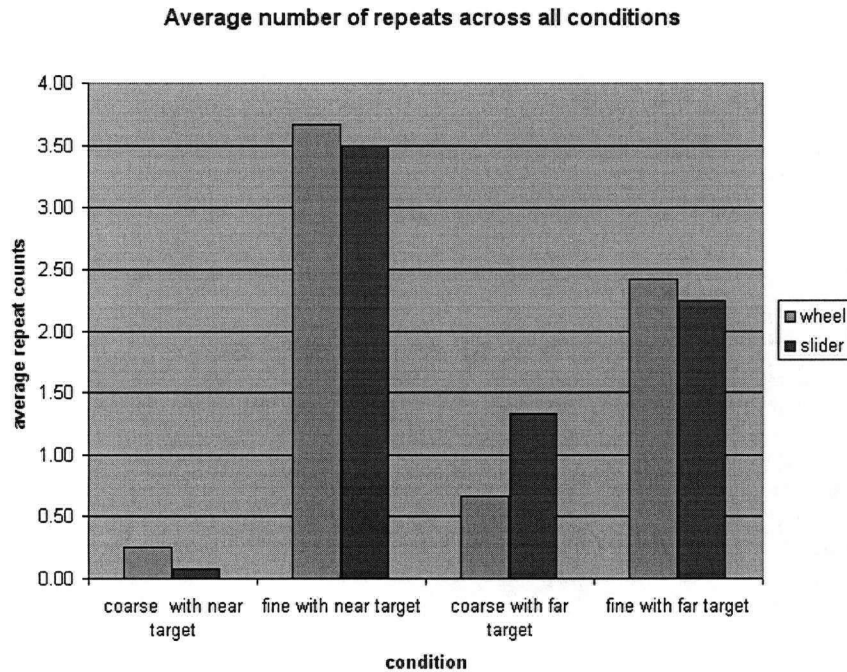
Figure 8.8: Amount of repetitive visits by subjects per trial across all eight conditions of Pilot 3.

ing category. The idea is to see if most of the subjects' moves fall within the refinement phase or fall within the approaching phase. By looking into this, we can find out if the interface is better suited for refinement or approaching the face targets.

Tables 8.7 and 8.8 show the means and standard deviations of the number of face locations visited that fall within refinement and approaching regions respectively. Figure 8.9 and 8.10 show the break down in percentage of the number of face locations visited during refinement and approaching stage. Notice that across all conditions, the slider interface always exceeds the wheel interface at

Wheel	Near Targets	Far Targets
Coarse Resolution	4.83	5.58
	(1.80)	(2.19)
Fine Resolution	7.75	6.25
	(5.59)	(3.77)
Dynamic Slider	Near Targets	Far Targets
Coarse Resolution	3.75	6.08
	(0.97)	(4.44)
Fine Resolution	11.83	10.33
	(8.64)	(5.19)

Table 8.7: Mean and standard deviation of the number of face locations visited in refinement region in Pilot 3.

Wheel	Near Targets	Far Targets
Coarse Resolution	1.17	6.17
	(2.82)	(3.86)
Fine Resolution	11.42	17.58
	(10.12)	(9.68)
Dynamic Slider	Near Targets	Far Targets
Coarse Resolution	0.42	3.92
	(0.51)	(4.98)
Fine Resolution	6.08	6.17
	(3.78)	(1.90)

Table 8.8: Mean and standard deviation of the number of face locations visited during approaching stage in Pilot 3.

the refinement region (see Figure 8.9). However, it is less than the wheel interface for the approaching region across all conditions (see Figure 8.10). This trend seems to suggest that the dynamic slider interface is better-suited for finding the neighbourhood, while the wheel interface is better suited for refining face matching because subjects travel less. The percentage data used in Figures 8.9 and 8.10 are listed in Table 8.9, under column headings "refine" and "approach" respectively.

ANOVA Table B.8 shows interface is not significant (F(1, 5)=2.20, MS_e=39.29,

Percentage of faces visited during refine phase
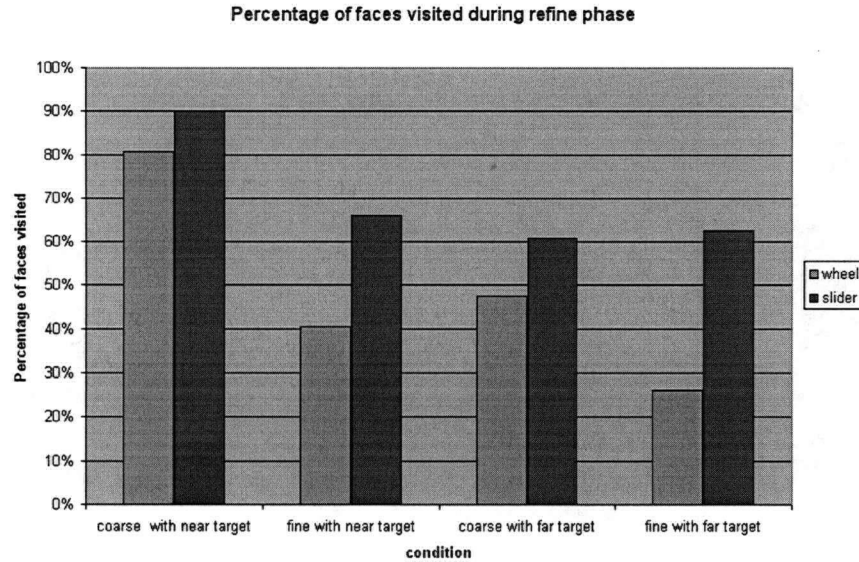
Figure 8.9: The histogram above shows a higher percentage of faces visited via sliders during the refinement stage of face navigation in Pilot 3.

$p > 0.05$); thus, we cannot reject H_{05} for Pilot 3. Resolution is a significant variable ($F(1, 5)=19.11, MS_e=19.89$, $p < 0.05$); the finer the resolution, the more refinement face locations visited. Target is insignificant ($F(1, 5)=0$, $MS_e=11.14$, $p > 0.05$). As for the interaction between the variables, Interface $\times$ Resolution is significant ($F(1, 5)=10.92$, $MS_e=10.54$, $p < 0.05$). We compute the simple main effect analysis in Table B.9 to study the interaction.

		Refine		Approach	
		Wheel	Slider	Wheel	Slider
Fine	Near	40	66	60	34
	Far	26	63	74	37
Coarse	Near	81	90	19	10
	Far	48	61	52	39

Table 8.9: Percentage distribution of face locations visited within refinement and approaching regions via the dynamic slider interface and the wheel interface (from Pilot 3).

Percentage of faces visited during approaching phase
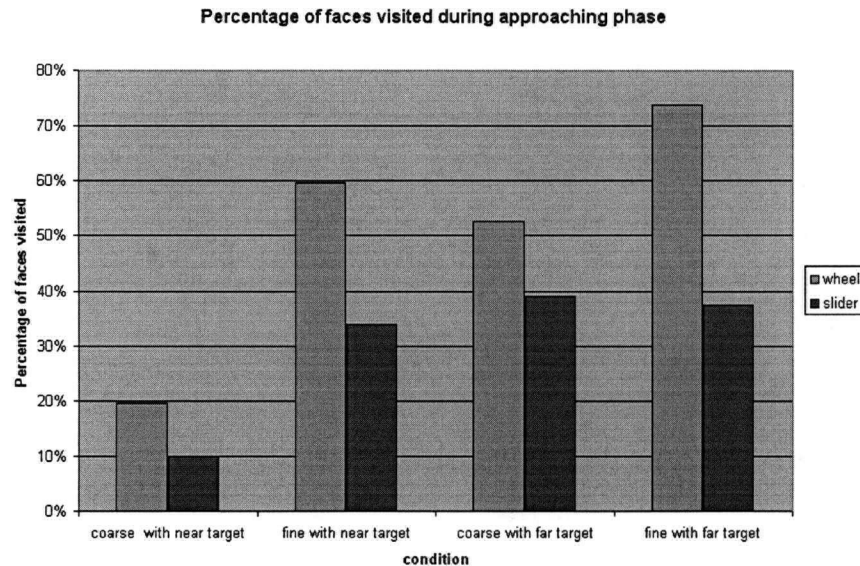
Figure 8.10: The histogram above shows that there is a higher percentage of face locations visited via the wheel interface in the approaching region of face navigation in Pilot 3.

According to Roberts et al. [50], if only two-way interactions are significant in a three-way ANOVA test, then simple main effect analysis is done with the two significant factors while ignoring the third. Cohen [15] suggests this be done by collapsing the third variable, i.e. averaging. In other words, since we are investigating the interaction between interface and resolution, the cells—which share the same interface and resolution condition but different target condition—are combined and averaged.

Table 8.10 shows the cell means after collapsing the target variable. Figure 8.11 shows the interaction plot. Our simple main effect in Table B.9 indicates that interface for the coarse resolution is not significant ($F(1,5)=0.03$, $p>0.05$);

Resolution Type

		Coarse	Fine
Interface	**Wheel**	5.21	7.00
	Dynamic Slider	4.91	11.08

Table 8.10: Cell means of refinement data after collapsing target variable in Pilot 3.

that is, there is no significant difference between cell means 5.21 and 4.91 (see Table 8.10). Interface for fine resolution is also insignificant ($F(1, 5)=5.09$, $p>0.05$); that is, there is no significant difference between cell means 7.00 and 11.08. Resolution for the wheel interface is not significant ($F(1, 5)=1.94$, $p>0.05$); that is, there is no significant difference between 5.21 and 7.00. However, resolution for slider interface is significant ($F(1, 5)=22.95$, $p<0.01$). In other words, in using dynamic sliders, subjects visit significantly more number of spots that are within two steps from the target.

As for the number of face locations subjects visited that fall within the approaching region, ANOVA Table B.10 indicates interface to be a significant source of variation ($F(1, 5)=7.84$, $MS_e=74.67$, $p<0.05$). It means that the wheel interface requires more steps than the slider interface in the approaching phase. This is not surprising because the dynamic sliders allow subjects to move multiple steps at a time while the wheel interface only allows subjects to move one step at a time. This allows us to reject H_{06} for Pilot 3. Resolution is significant ($F(1, 5)=36.63$, $MS_e=35.84$, $p<0.01$) indicating the finer the resolution, the more steps it takes to get to the target. Target is significant ($F(1, 5)=12.74$, $MS_e=0.02$, $p<0.05$) indicating the further the target, the more steps required to get to there.

Interaction between interface and resolution for refinement phase
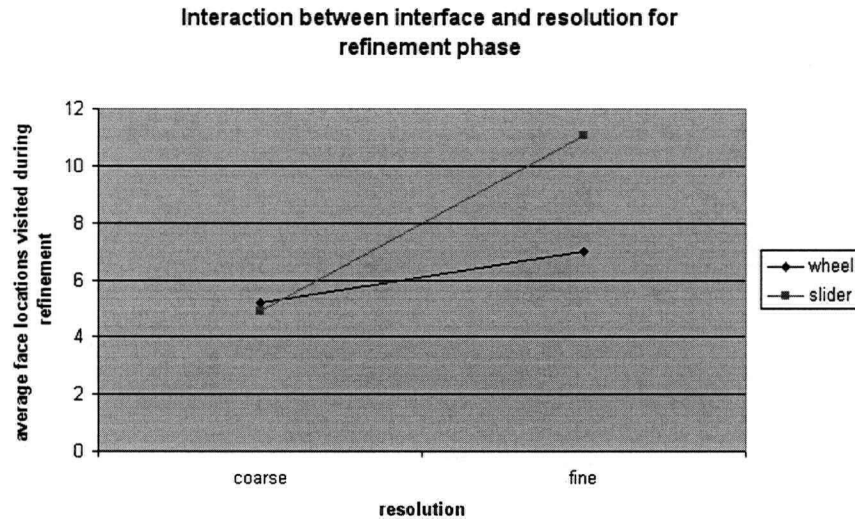
Figure 8.11: Plot showing the interaction between resolution and interface type for the refinement data.

The only interaction that is significant is that of resolution and interface $(F(1, 5)=7.73, p<0.05)$. Table 8.11 shows the cell means after collapsing the target variable. Figure 8.12 shows the interaction plot. Our simple main effect in Table B.9 indicates that interface for the coarse resolution is not significant $(F(1,5)=0.36, p>0.05)$; that is, there is no significant difference between cell means 3.67 and 2.17. Interface for fine resolution is significant $(F(1, 5)=11.27, p<0.05)$; that is, subjects step through fewer face locations using the dynamic sliders than using the wheel. Resolution for the wheel interface is significant $(F(1, 5)=39.30, p<0.01)$; that is, subjects take a greater number of steps to approach the two step neighbourhood via the wheel interface under fine resolution. Resolution for the slider interface is insignificant $(F(1, 5)=5.25, p>0.05)$; that is, there is no significant difference between cell means 2.17 and 6.13.

		Resolution Type	
		Coarse	**Fine**
Interface	**Wheel**	3.67	14.50
	Dynamic Slider	2.17	6.13

Table 8.11: Cell means of approaching phase data after collapsing target variable in Pilot 3.

Interaction betweeen interface and resolution for approaching phase
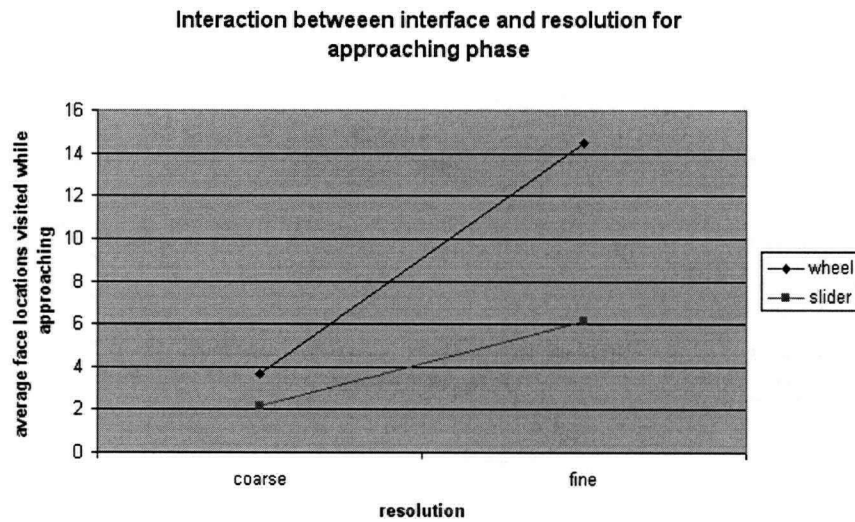
Figure 8.12: Plot showing the interaction between resolution and interface type for the approaching phase data.

In order to ensure that our results are not biased, we have also checked our data for learning effects. We do not detect this in our results.

The thing to note from Pilot 3 reported statistics is that subjects appear to do poorer under fine resolution, when looking for near targets. This is counter intuitive as near targets require less travelling than far targets. We also notice that the majority of the face locations visited by subjects via sliders fall within refinement, while the majority of the face locations visited by subjects via wheel

fall within approaching phase. This suggests that in order to minimize the number of visits, sliders may be better for reaching the neighbourhood and the wheel interface may be better for the refinement phase. However, it must be noted that in Pilot 3, for coarse resolution the actual number of visits during the refinement phase was slightly lower using sliders than the wheel interface.

8.3.5 Interview

In Pilot 3, all subjects find it more demanding when navigating under fine resolution. One subject remarks that he spends a lot of time trying to get the skin tone right. Some subjects comment that they lean forward to get a closer look at the faces when the resolution appears to be fine. Another subject mentions that he relies on the wheel's "back" function a lot when he is guessing where to go. Although a couple of subjects still prefer the wheel interface, most subjects prefer the slider interface because they have more sense of where to go. All subjects reported that they prefer coarse resolution over fine resolution.

8.4 Discussion

For Pilot 3, we will go over the ANOVA results by referring to the "Pilot 3" rows in the ANOVA summary Table B.1. Note that the interface factor only shows significance during the approaching phase, hence the only hypothesis rejected is H_{06} (see Table B.2). We discuss each factor in the order of resolution, interface (and its interaction with resolution), and target.

8.4.1 Effect of Resolution

As observed from the ANOVA summary (Table B.1), resolution is a frequent factor that contributes to significant variation in Pilot 3. It is only insignificant when the dependent measurement is the subjects' scores. From our observations, resolution seems to be the key that turns the wheel interface from useful to useless.

Subjects find it particularly hard to use the wheel interface under the fine resolution condition. In fact, fine resolution induces more trial-and-error behaviour because local face neighbours are not helpful to inform subjects which direction to go. This is when the face gradients fail to assist users. The trial-and-error behaviour is most apparent with nearby targets (see Figure 8.8).

Dynamic sliders, on the other hand, provide faces at the extreme values with respect to the current face. Extreme faces are not really affected by resolution, be it coarse or fine because the boundary of the face space remains the same. Although the dynamic slider interface does not display the immediate face gradients, subjects can estimate the gradients by observing the extreme faces. This proves to be more useful than the wheel interface which provides subjects with local neighbours.

Since subjects find it particularly challenging to detect the subtle facial differences when using the wheel interface under fine resolution, it would be very helpful to be able to toggle the current face with the target face at the centre of the wheel. It would also be very helpful if users could drag the target face around for direct juxtapositions with the neighbouring faces.

8.4.2 Effect of Interface

From the result summary, ANOVA Table B.1, the interface variable is only significant when the dependent measure is the number of face locations visited in the approaching phase. Its interaction with resolution is significant when the measurements are face location counts in both refinement and approaching phases. We investigate the interactions by computing the simple main effect tests. We find that it is really due to the fact that the dynamic sliders have the advantage of moving multiple steps at a time whereas the wheel interface does not. Nevertheless, subjects most likely have various strategies for maneuvering in the face space prior to reaching the region of close proximity, and it would be an interesting future study to observe people's converging methods by varying the starting points of navigation in the face space.

We observe that subjects visit a fewer number of face locations which fall in the refinement region when navigating with the wheel interface than the dynamic sliders (see Figure 8.9). On the other hand, subjects visit a fewer number of face locations that fall in the approaching region when using the dynamic sliders in comparison to the wheel (see Figure 8.10). The implication that the wheel interface is more effective for face match refinement, while the slider interface is more helpful for reaching the neighbourhood, is therefore plausible. This trend persists in the final experiment.

8.4.3 Effect of Target

From the result summary, ANOVA Table B.1, the target variable is only significant when the dependent measurement is the face location counts in approaching phase. As we find with ANOVA, the results indicate the trend that the further the target, the more steps required to approach the neighbourhood.

One interesting trend is that subjects have the most trouble searching for near targets while under fine resolution conditions (Figure 8.7). This is especially apparent when they use the wheel interface. Such a result is counter-intuitive as we would expect further targets to be more challenging since subjects are required to manipulate more axes. Nevertheless, this may be due to the fact that targets closer to the original are less prominent, and therefore, more difficult to spot. This trend, however, is not apparent in the final experiment with a larger group of subjects.

8.4.4 Effect of Skin Tones

As a sanity check, we notice that the condition of coarse resolution with far target is really similar to the Pilot 2 condition of uncorrelated axes. Resolution, axes type and face target positions are exactly the same, but the choice of principal components is different. This may explain the difference between the average number of revisits between these two testing conditions. Table 8.12 shows the average repetition counts and note the large difference between 6.58 and 0.67 revisits for the wheel and 5.25 and 1.33 revisits for the dynamic sliders. What may have caused this difference?

	Pilot 2	Pilot 3
Resolution	coarse	coarse
Axes Type	uncorrelated	uncorrelated
Target Type	$\{0,1,1,1,2,2\}$	$\{0,1,1,2,2\}$
Wheel	6.58	0.67
Dynamic Slider	5.25	1.33

Table 8.12: Average revisit counts of the uncorrelated axes condition of Pilot 2, versus the coarse resolution with far target condition of Pilot 3. Note that both conditions have similar resolution, axes type and target type conditions. (We show the set of numbers that form the targets in both apparatus). However, the average amount of repetition between the two conditions is large. This may be due to the choice of texture principal components.

Since the only difference in apparatus is the choice of principal components, we see that the texture navigation axes used in Pilot 2 may not be as distinct in skin tones for subjects. Figures 8.13 and 8.14 show the facial textures used in Pilot 2 and 3 respectively. Note that Pilot 2 uses two texture axes that moves from pink skin tones to green skin tones (the top two sliders of Figure 8.13) and this can be potentially confusing to subjects if they rely heavily on differentiating faces by skin colours. The texture axes used in Pilot 3 have three very distinct skin tones (see Figure 8.14) in contrast, and should therefore be easier to differentiate.

We have provided an account of our Pilot 3 experiment. The formal experiment is outlined in the next chapter.
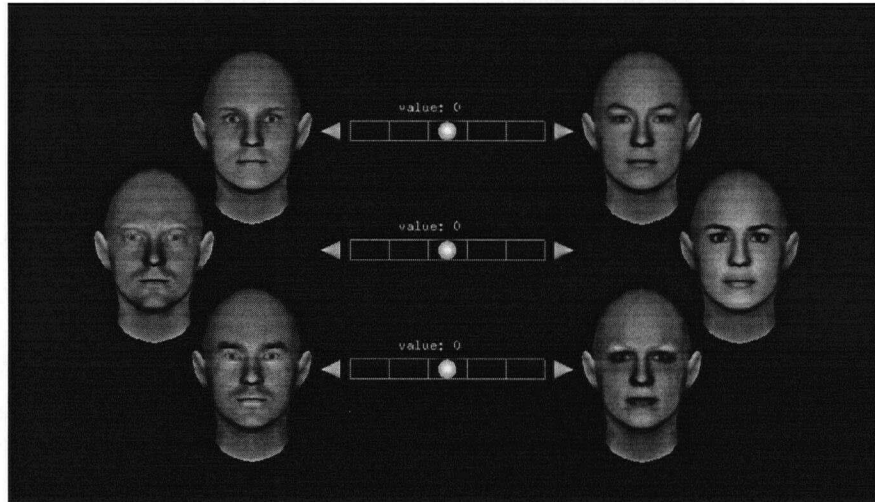
Figure 8.13: Texture navigation axes (uncorrelated) of Pilot 2. Note that the top texture axis allows choices between greenish skin tones to pinkish skin tones. The middle axis goes from pink skins to greenish skins and the bottom navigation axis is from orange beige to gray beige.
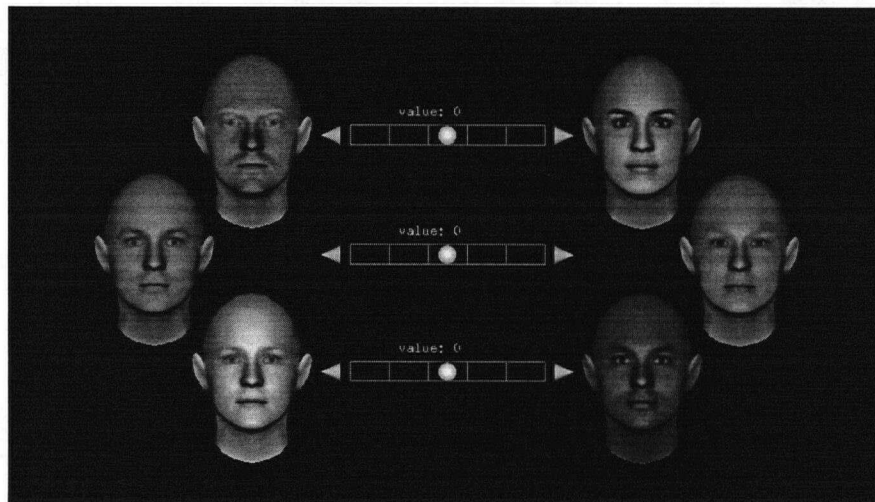
Figure 8.14: Uncorrelated texture navigation axes of Pilot 3. Note that the top texture axis allows navigation from pink skin tones to green skin tones. The middle axis goes from blue to yellow skin tones and the bottom texture axis goes from white to brown skin tones.

Chapter 9

The Formal Experiment: Effect of Interface, Resolution and Target Distance

9.1 Introduction

Given the intriguing findings from Pilot 3, we decide to do a formal experiment using the same experimental conditions. The independent variables are interface, resolution and target type. Table 8.1 shows the experimental condition for the two interfaces we test.

9.2 Method

Since the experimental conditions are like those of Pilot 3, we do not repeat the descriptions of apparatus and procedure here. Thus, these sections are omitted.

9.2.1 Subject

In this experiment, we recruited fifteen volunteers as subjects, of whom eight are computer science students, two are engineering students and the rest have other backgrounds and professions. There are seven men and eight women; of

whom three are Caucasian, two are East Indian and nine are Oriental.

9.3 Results

We test for the same hypotheses as in Pilot 3:

- H_{01}: Subjects take a similar amount of time using the wheel interface and the slider interface.

- H_{02}: Subjects score similarly using the wheel interface and the slider interface.

- H_{03}: Subjects revisit faces a similar number of times using the wheel interface and the slider interface.

- H_{05}: Subjects visit a similar number of faces during the "refinement phase" of face matching tasks, using both the wheel interface and the slider interface.

- H_{06}: Subjects visit a similar number of faces during the "approaching phase" of face matching, using both the wheel interface and the slider interfaces.

We report statistical results in the same order as the hypotheses but group H_{05} and H_{06} tests together, as they are closely related. An interview summary is included at the end of this Result Section.

9.3.1 Performance in Time

Figure 9.1 show the average time subjects take to complete the trials. The wheel interface takes a bit more time overall, except in the case of coarse resolution used with near targets (column one of the histogram). The cell means and standard deviations are shown in Table 9.1.

Average time taken per trial across conditions
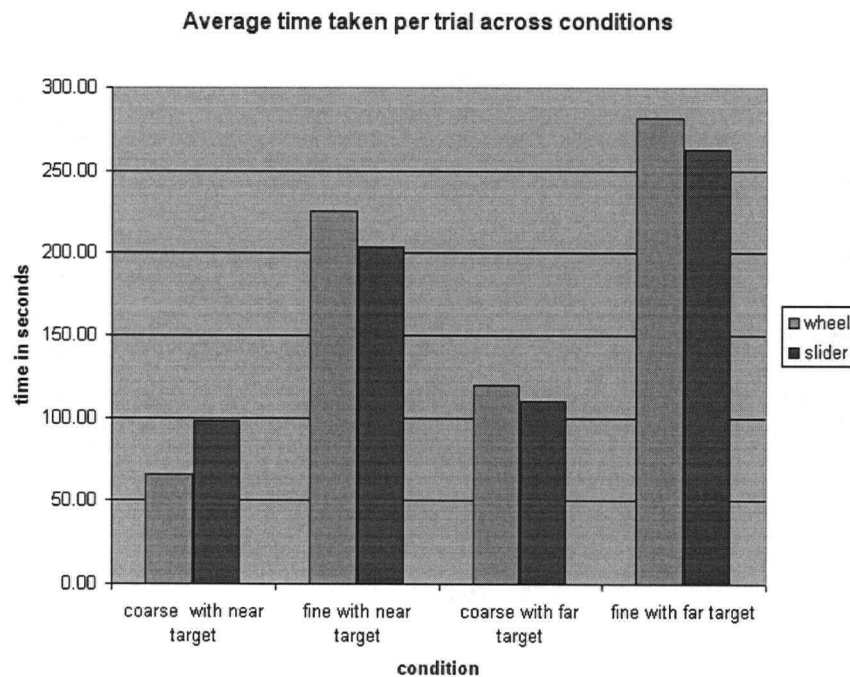
Figure 9.1: Average time taken by subjects across all eight conditions in the last experiment.

ANOVA test shown in Table B.12 does not indicate interface to be a significant variable (F(1,5)=0.21, MS_e=5528.40, p>0.05); thus, we do not have sufficient evidence to reject H_{01} for this experiment. Resolution, however, is significant (F(1, 5)=88.86, MS_e=14189.80, p<0.01). This shows that the finer

Wheel	Near Targets	Far Targets
Coarse Resolution	65.58s (42.62s)	119.87s (65.93s)
Fine Resolution	225.49s (180.98s)	281.82s (82.26s)
Dynamic Slider	**Near Targets**	**Far Targets**
Coarse Resolution	98.26s (97.69s)	110.43s (65.31s)
Fine Resolution	203.54s (140.34s)	263.14s (81.49s)

Table 9.1: Mean and standard deviation of the time taken by subjects in the final experiment.

the resolution, the longer subjects need to reach the targets. Target is a significant variable (F(1, 5)=8.21, MS_e=15189.90, p<0.05); that is, the further the target, the more time is required. Lastly, none of the two-way and three-way interactions are significant.

9.3.2 Performance in Score

Subjects' scores are shown in Figure 9.2. Subjects appear to do slightly better using the wheel interface overall, except for when they use fine resolution for navigating to far targets (column seven of the histogram). This result is slightly different from our pilot test (Figure 8.7) where subjects seem to do slightly better using sliders overall. Table 9.2 shows the cell means and standard deviations.

ANOVA Table B.13 indicates the interface factor is not significant (F(1, 5)=0.00, MS_e=0.05, p>0.05); therefore, we cannot reject H_{02} for the final experiment. Resolution is a significant variable (F(1, 5)=16.87, MS_e=0.08, p<0.01); that is, the finer the resolution, the lower the score. Target is not

Average score across conditions
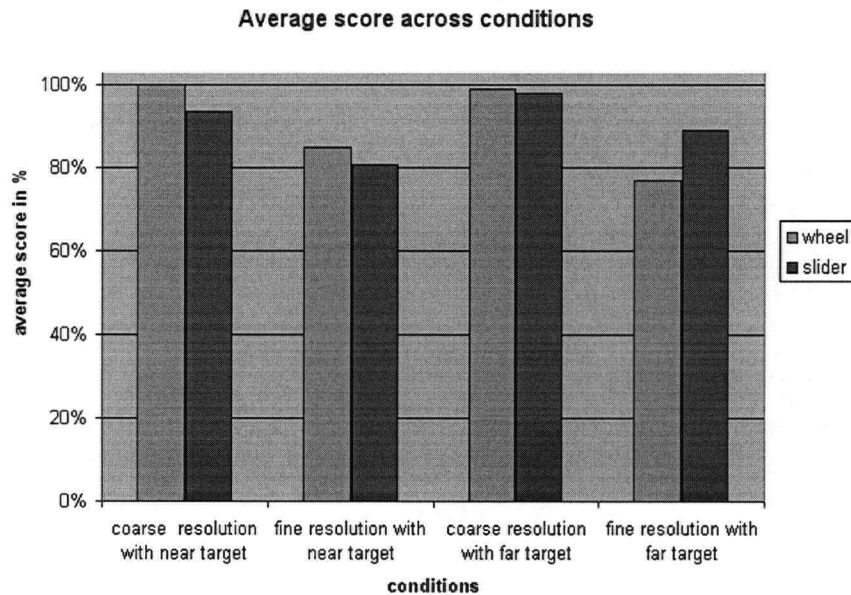
Figure 9.2: Accuracy score of subjects across all eight conditions in the last experiment.

significant (F(1, 5)=0.23, MS_e=0.03, p>0.64). Finally, no interactions between the variables are detected to be significant.

9.3.3 Performance in Reducing Repetition

We also accumulate counts of the number of revisits of faces in this experiment. The break down is shown in Figure 9.3. This figure is quite different from our Pilot 3 experiment, shown in Figure 8.8. Subjects, on average, appear to revisit faces much more often than the Pilot 3 subjects when using the wheel interface—especially under the fine resolution condition. The cell means and the standard deviations are given in Table 9.3.

In our ANOVA test of the repetitive measurements (see Table B.14), we find

Wheel	Near Targets	Far Targets
Coarse Resolution	100% (0%)	99% (5%)
Fine Resolution	85% (26%)	77% (28%)
Dynamic Slider	**Near Targets**	**Far Targets**
Coarse Resolution	93% (21%)	98% (8%)
Fine Resolution	81% (33%)	89% (18%)

Table 9.2: Mean and standard deviation of subjects' scores in the final experiment.

Wheel	Near Targets	Far Targets
Coarse Resolution	1.4 (2.39)	2.43 (3.31)
Fine Resolution	6.13 (8.34)	7.33 (6.96)
Dynamic Slider	**Near Targets**	**Far Targets**
Coarse Resolution	2.10 (3.76s)	1.67 (3.93)
Fine Resolution	3.63 (4.46)	3.87 (3.22)

Table 9.3: Mean and standard deviation of the number of revisits in face space from the final experiment.

Repeat counts across conditions
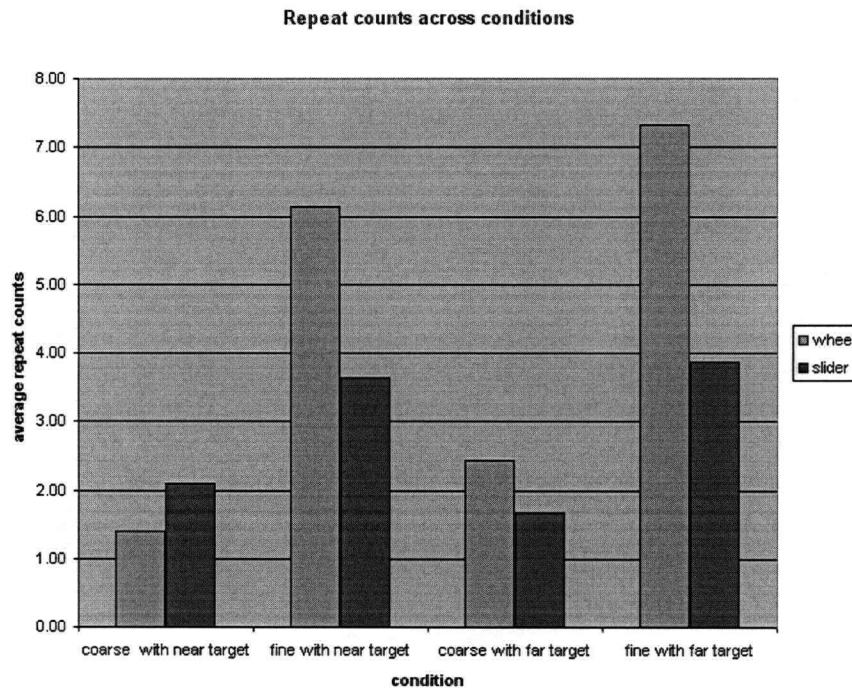
Figure 9.3: Revisit counts for subjects across all eight conditions in the last experiment.

interface to be a significant factor (F$(1, 5)$=5.28, MS_e=25.84, p<0.05); that is, subjects using the wheel interface seem to have more repetition especially when the resolution is fine. This suggests that different interfaces cope with the resolution condition differently. Not surprisingly, we also find resolution to be significant (F$(1, 5)$=28.72, MS_e=23.33, p<0.01) affirming that the finer the resolution, the more repetitions. We observed that subjects find the wheel interface to be particularly difficult to use under the fine resolution condition. This is likely due to the fact that face prominence is below the "just noticeable difference" threshold. Thus, having gradient information at local regions is

Resolution Type

		Coarse	Fine
Interface	**Wheel**	1.92	6.73
	Dynamic Slider	1.88	3.75

Table 9.4: Cell means of approaching phase data after collapsing target variable in Pilot 3.

not very helpful to them. Target is not a significant variable ($F(1, 5)=0.8$, $MS_e=19.42$, $p>0.05$) in this case.

The only interaction that is significant is resolution $\times$ interface ($F(1,5)=7.08$, $MS_e=018.43$, $p<0.05$). Table 9.4 shows the cell means after collapsing the target variable. Figure 9.4 shows the interaction plot. Our simple main effect in Table B.15 shows that interface for the coarse resolution is not significant ($F(1,5)=0.00$, $p>0.05$); that is, there is no significant difference between cell means 1.91 and 1.88. Interface for fine resolution is not significant ($F(1, 5)=3.69$, $p>0.05$); that is, there is no significant difference between cell mean 6.73 and 3.75. Resolution for the wheel interface is significant ($F(1, 5)=13.36$, $p<0.05$); that is, when using the wheel interface, subjects revisit more frequently when the resolution is fine. Resolution for the slider interface is insignificant ($F(1, 5)=2.01$, $p>0.05$); that is, there is no significant difference between cell means 1.88 and 3.75.

We further check if subjects' scores correlate with the number of revisits. Figure 9.5 shows that they are negatively correlated. This implies that subjects who do well in face matching tasks tend to know how to use both interfaces effectively, thus reducing the number of repeat counts.

Interaction between interface and resolution in repetition
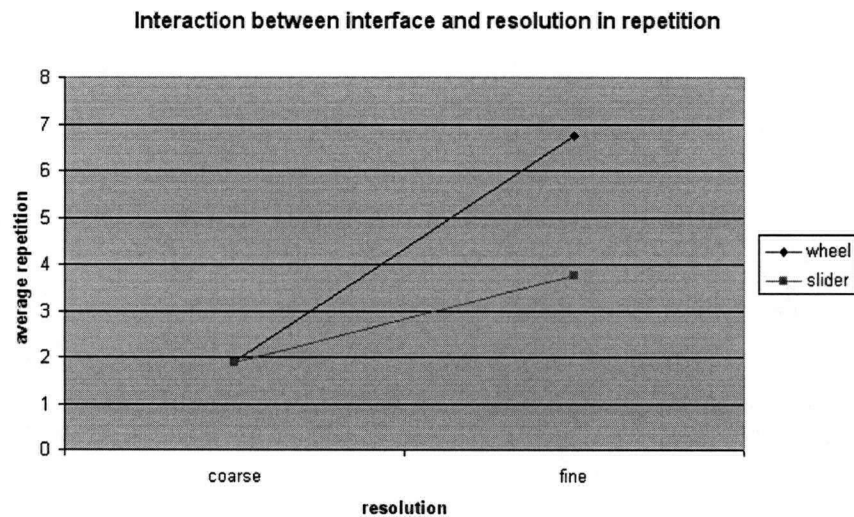
Figure 9.4: Plot showing the interaction between resolution and interface type for the repetition data.

9.3.4 Performance in Refinement and Approaching

We further analyze the data in the refinement and the approaching regions. The cell means and standard deviations are given in Tables 9.5 and 9.6.

Just as in Pilot 3, we calculate the percentage of faces visited in both the refinement and the approaching regions. We again notice that when subjects use the slider interface to navigate, they visit a greater number of places in the refinement region (Figure 9.6). However, when they use the wheel interface, they visit a greater number of positions that fall in the approaching regions (see Figure 9.7). These are the trends we expect, which implies that the slider interface is suitable for reaching the neighbourhood, while the wheel interface is better for refining. The percentage data used in Figures 9.7 and 9.6 are listed

Wheel	Near Targets	Far Targets
Coarse Resolution	5.5 (3.28)	6.27 (3.61)
Fine Resolution	12.23 (12.34)	8.3 (8.69)
Dynamic Slider	**Near Targets**	**Far Targets**
Coarse Resolution	7.60 (7.90)	7.10 (4.96)
Fine Resolution	10.20 (9.32)	10.90 (4.74)

Table 9.5: Mean and standard deviation of the number of face locations visited which fall in the refinement region in the final experiment.

Wheel	Near Targets	Far Targets
Coarse Resolution	4.07 (11.37)	7.00 (7.03)
Fine Resolution	9.67 (11.27)	19.83 (9.32)
Dynamic Slider	**Near Targets**	**Far Targets**
Coarse Resolution	1.60 (4.48)	4.00 (4.75)
Fine Resolution	7.33 (6.13)	12.07 (7.99)

Table 9.6: Mean and standard deviation of the number of face locations visited which fall in the approaching region in the final experiment.

Score vs. revisits across subjects
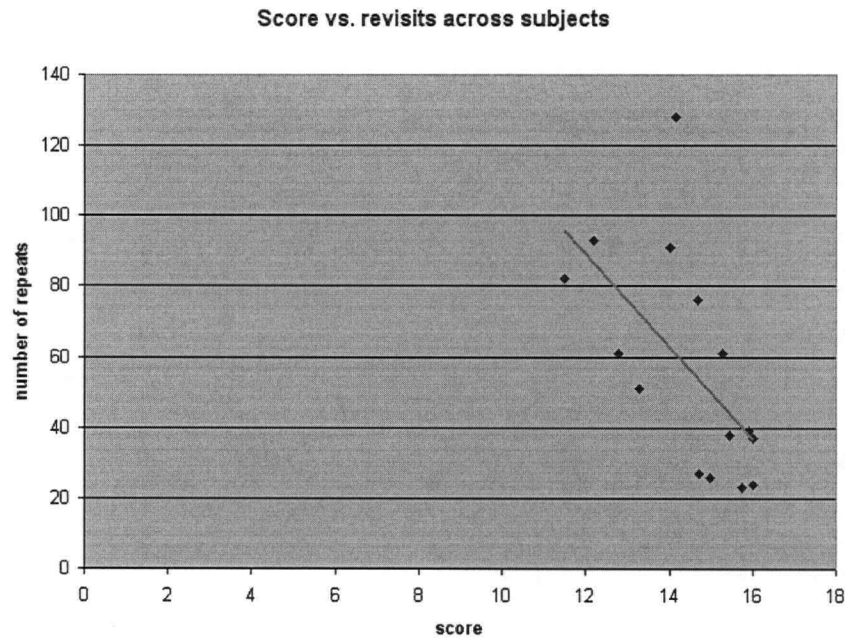
Figure 9.5: Correlation between subjects' performance and face revisit counts.

in Table 9.7, in the "refine" column, and the "approach" column respectively. The trend is similar to the ones found in Pilot 3 (Figures 8.10 and 8.9).

ANOVA test (see Table B.16) reveals that when the dependent measure is the face locations visited in refinement, interface is not a significant variable ($F(1,5)=1.95$, $MS_e=23.50$, p>0.05). This means we cannot reject H_{05} for the final experiment. Resolution is a significant variable ($F(1, 5)=34.18$, $MS_e=25.24$, p<0.01). This indicates that the finer the resolution, the more face locations to step through. Target is not significant ($F(1,5)=0.31$, $MS_e=106.30$, p>0.05). Lastly, none of the interactions between the variables are significant.

As for the approaching phase, ANOVA test Table B.17 reveals that inter-

Percentage of faces visited during approach phase
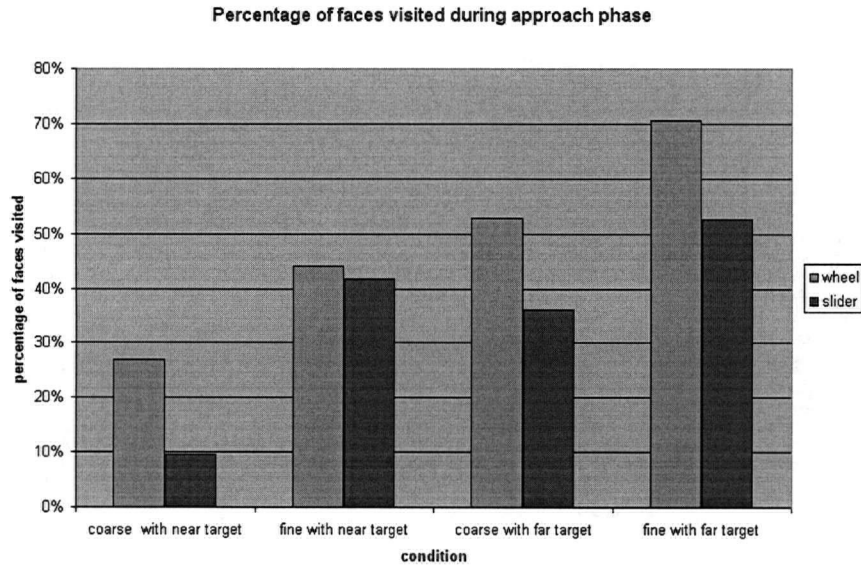
Figure 9.6: The histogram above shows there is a higher percentage of faces visited via the wheel interface during the approaching stage of face navigation in the last experiment.

face is significant (F(1, 5)=16.81, MS_e=54.05, p<0.01). We thus have enough evidence to reject H_{06} for the final experiment. The statistics indicate that subjects had to step through more face locations using the wheel interface than the slider interface before reaching the two-step neighbourhood. As mentioned before, this is expected as the wheel interface can only move one step at a time.

		Refine		Approach	
		Wheel	Slider	Wheel	Slider
Fine	Near	56	58	44	42
	Far	30	47	70	53
Coarse	Near	73	90	27	10
	Far	47	64	53	36

Table 9.7: Percentage distribution of face positions visited that fall in the refinement region and the approaching region via the dynamic sliders and the wheel interface (from the final experiment).

Percentage of faces visited during refine phase
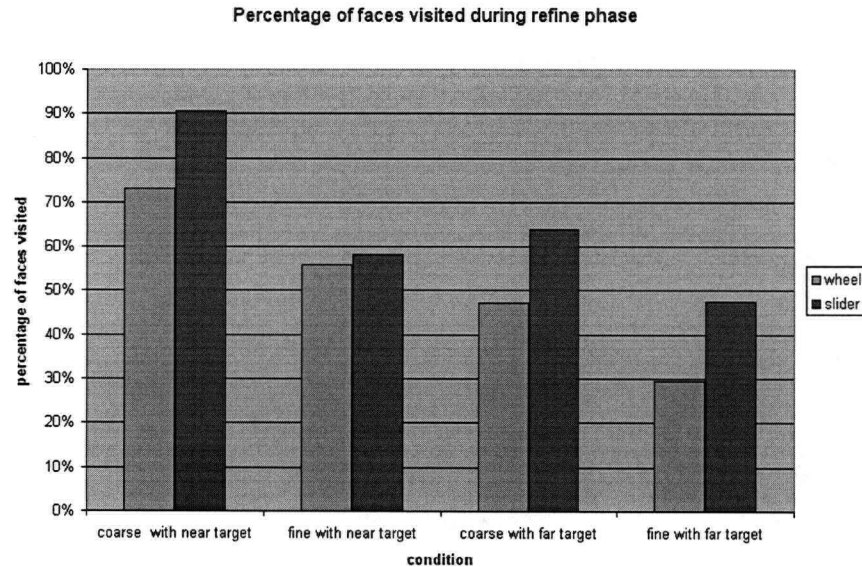
Figure 9.7: The histogram above shows a higher percentage of faces visited via sliders during the refinement stage of face navigation in the last experiment.

Resolution is significant $(F(1, 5)=34.96, MS_e=112.30, p<0.01)$. The finer the resolution, the more number of steps are required to get through. Target is also significant $(F(1,5)=24.04, MS_e=63.86, p<0.01)$. This shows that the further away the target, the more steps subjects generally take to reach the neighbourhood. As for the interactions between the variables, none of them is significant.

Similar to previous experiments, we have checked our data. We do not detect any significant learning effect.

The noteworthy observation from this experiment is that subjects' scores and their repetitiveness in navigation are negatively correlated. Therefore, subjects who score higher tend to have fewer repetitions. Like Pilot 3, subjects also

visit fewer faces using the wheel interface for refinement when compared to the dynamic sliders and visit fewer faces using dynamic sliders to get to the bulk part when compared to the wheel. This highlights the fundamental difference in their design.

9.3.5 Interview

In this experiment most subjects find user testing to be a compelling experience. All subjects prefer the dynamic slider interface because the end faces provide better visual clues for them to work out where to navigate. Subjects can see what faces to expect. One subject comments that she prefers the sliders because the faces are arranged linearly and she is used to reading from left to right. One subject also remarks that he finds the diagonal arrangement of opposite faces to be new. He would prefer the wheel interface to run faster so that he can press down the mouse button and preview the faces at the extreme values. Another subject even suggests that we label the axes with meaningful terms which is precisely what we want to avoid by designing configural approach and navigating via principal components! As for resolution, all subjects prefer coarse over fine.

9.4 Discussion

For this experiment, we will go over the ANOVA results by referring to the "Final Experiment" rows in the ANOVA summary (Table B.1). Note that the interface factor is only significant when the dependent measurement is repetition

or face location counts in the approaching phase. This is indicated by the rejection of H_{03} and H_{06} (see summary in Table B.2). We discuss each factor in the order of resolution, interface (and its interaction with resolution) and target type.

9.4.1 Effect of Resolution

Just like in Pilot 3, resolution is a variable that turns up significant persistently as shown in the ANOVA summary (Table B.1). Having the appropriate resolution is thus vital. The important thing to be aware of with resolution is: when it goes below the subjects' "just noticeable difference" level for faces, the wheel interface is not effective. As one subject remarks, he wishes there were labels on the wheel axes so that he can see where he is going. Another subject's comment spotlights this flaw as well. He suggests that if the faces can be generated faster, then he can potentially press down the left mouse button on a particular face to see its extreme values. In an actual system, resolution would be an adjustable parameter as the face space is continuous; therefore this idea may work well.

Sliders, on the other hand, over come this problem by showing the extreme faces. People can estimate the face gradients between the extreme faces, although this comes at the cost of trial-and-error by guessing the face gradients in between. This shows that it is very helpful to provide the global direction of the face space so that users can recognize where to go, rather than having to recall where to explore.

9.4.2 Effect of Interface

The interface variable is significant when the dependent measurement is repetition counts as observed from the ANOVA summary (Table B.1). Compared to Pilot 3, subjects have more average repetitions in the final experiment (see Figure 8.8 for Pilot.3 and Figure 9.3 for the final experiment). This may be due to a wider sample of subjects.

We also see that the higher subjects score, the better they are at avoiding repetitions (see Figure 9.5). This is a sign that indicates our subjects likely have different levels of face recognition skills. We know from our results that under fine resolution, the wheel interface breaks down. Subjects cannot comfortably distinguish all the faces; thus from guessing, they revisit more faces (see Figure 9.3). If subjects who score higher have fewer repetitions, then they must have sharper face recognition abilities.

Interface ·× Resolution is also significant (as Table B.1 indicates). Unlike the results in Pilot 3, subjects do not have as many revisits when retrieving nearby face targets under a fine resolution using both interfaces (compare column three and four of Figures 8.8 and 9.3). What is consistent in both experiments, however, is that subjects appear to "wander" less using the wheel interface during face refinement when compared with sliders. On the other hand, they approach the neighbourhood more efficiently when using the slider interface as Figures 9.7 and 9.6 indicate. This tells us that perhaps people do not need the gradient info at the beginning of navigation when all they need to do is to get into the correct zone by using sliders. Once they are close enough, gradient info

can help with refinement and thus help subjects reach the target. However, if the resolution is too fine, then it is useful to implement mechanisms that allow side-by-side comparisons.

From ANOVA Table B.1, interface is again significant when the dependent measurement is the number of face locations visited in approaching. We learn from our statistical results that this is really due to the fact that the wheel interface only allows subjects to move one step at a time instead of multiple steps like dynamic sliders.

9.4.3 Effect of Target

From ANOVA Table B.1, the target variable is significant when the performance is measured in time, as well as when the performance is measured in the number of face locations visited during approaching. This simply indicates that the further away the targets, the longer the time is needed to navigate to targets (see Table 9.1). More steps are also required to approach close proximity (see Table 9.6). This is expected because the wheel interface requires more button clicks when compared with the sliders. Although Pilot 3 predicts subjects are likely to find further targets to be more distinguishable and thus "easier" to reach, this trend is not apparent here.

In this chapter, we provide an account of the last experiment conducted. We have completed all descriptions of experiments. Summaries of hypotheses testing are in Table 5.1 in Appendix B. Our conclusion chapter is next in which we tie together loose ends.

Chapter 10

Conclusion and Future Work

In the previous chapter, we present results from experiments which test a number of interfaces inspired by the art and folklore of the sculptures of Sanjusangendo. The interface engages users during the navigation process and is similar to many works by Fels [22] which extend the human body and mind. Users need to interact with the interface in order to navigate to a destination in the face space. This method not only depends heavily on users' face recognition skills but also on the face gradients presented. Our results are very interesting because people, as human beings, identify with faces so much that they are easily affected by them.

This is also reflected in research by [44] and [62]. However, our research is different in that it is not about interacting with artificial agents; it is about differentiating faces and combining features, forcing users to have an intimate relation with faces they are not familiar with. In fact, one of our subjects observes that the experiment is not so much about face recognition, but rather about performing addition and subtraction with faces. This interpretation is absolutely correct as we resonate Klawe's research [29] on educational games for children by making our experiments game-like.

10.1 Contributions

We have made a number of contributions. We initiate and explore a different point of view in investigating the face retrieval problem in a large face space. More specifically, we place emphasis on the retrieval process by means of navigation in our interface design. Our interfaces avoid the pitfalls of component retrieval systems by providing realistic and colourful morphed faces, staying clear of the verbal categorization of faces, and allowing users to navigate from the average face to distinctive ones.

The four experiments we conduct have interesting outcomes.

10.1.1 Pilot 1

In Pilot 1, our wheel interface, while not being significantly faster than the sliders in our tests, highlights a number of noteworthy findings. The wheel interface is definitely useful for providing users with gradients of faces for selection. Having the gradient information helps users concentrate on studying the faces rather than spending time clicking the mouse in trial-and-error uncertainty. Our test is not complicated enough for our data to show large variations over all conditions. However, we strongly believe that sufficient results are found to gather the benefits of each interface. Participants spend less time per face with the wheel interface and show direct convergence on the target. With static sliders, participants exhibit more trial-and-error behaviour, which is successful as long as the dimension of the space is not too large. Perhaps a hybrid system of both is the ideal approach. Finally, a majority of subjects are more comfortable with

the wheel interface, suggesting it offers a viable means for face navigation.

10.1.2 Pilot 2

In Pilot 2, our wheel interface demonstrates that it is more helpful to subjects navigating with correlated axes, which are not as intuitive to use as uncorrelated ones. Although the effect is not significantly large, it does point out a possible advantage in gradient-based interface: it may help users make better face selections, and is, thus, may be better for refinement.

10.1.3 Pilot 3 and the Last Experiment

In Pilot 3 and the last experiment, we illustrate that the wheel interface is more suitable for refinement; whereas, the slider interface is better suited for reaching the face neighbourhood. We also show how the gradient-based method can break down when the distinctiveness of faces is below the just noticeable difference. Thus, as long as the face gradients are distinguishable, the wheel interface is not impaired by fine resolution.

10.2 Lessons Learned

We have also learned a number of valuable lessons along the way.

10.2.1 People Like Faces

We learn that it is not always appropriate to apply user testing methods for colour matching to face matching because faces are more engaging than colours.

Subjects do not necessarily zoom quickly into the close proximity region of a target face, simply because they enjoy the process of arriving. In fact, we observe that most subjects find the test to be an eye-opening experience, particularly when they recognize faces that look like celebrities, friends or family members. Therefore, it is not appropriate to set a five minute time limit per trial as we did in Pilot 1 (we were simply following the example of Schwarz [52]).

10.2.2 Face Recognition Skills

We also find that subjects' different face recognition abilities is of interest. As Biederman mentions in [7], good face recognizers use the whole face, and poor face recognizers tend to pick a small set of distinctive features. Different face recognition skills relate to how well recognizers can discern faces under different degrees of resolutions. This may explain why some subjects fail to reach their targets even though they are just one step away. (In the case of the wheel interface, the target faces are right in front of the subjects!)

10.2.3 Fitts' Law for Face

It is also suggested that at a later stage in our research, it might be worthwhile to model the time it takes a user to reach a face target by modifying Fitts' Law[1]. This is of interest because the Fitts' Law relation between the distance

---

[1]Fitts' Law is a model to account for the time it takes to point at something, based on the size and distance of the target object. Fitts' Law and variations of it are used to model the time it takes to use a mouse and other input devices to click on objects on a screen. It is defined as follows: $MT = a + blog_2(\frac{A}{W} + c)$. The variable MT is the movement time, a and b are regression coefficients, A is the distance of movement from start to target, W is the width of the target, and c is a constant of 0.05 or 1 (see [36] for details).

of movement and the target size parallels the face navigation movement and the target face. Although the target face is fixed in size for ease of comparison, its prominence can affect the time users take to navigate there. Therefore, facial prominence can replace target size as a variable when we attempt to model a formula similar to Fitts' Law. It may also be a good idea to model the distance between the spectrum of face options and face targets to see if it effects users' comparisons. The difference between face recognition and object recognition, outlined by Biederman [7], however demands careful scrutiny between the difference in reaching objects and reaching faces on the screen. Nevertheless, there are potential values in doing this. In modelling the time it takes people to reach face targets, we can further improve navigation interfaces for faces.

10.3 Future Work

For future work, we want to investigate the effects of other parameters, such as the starting position of navigation and face matching with realistic face targets. We also want to explore methods of organizing the 128 principal component axes in meaningful ways to facilitate subjects in selecting the axes with the right features for navigation. As there is no widely accepted face similarity metric presently, we believe this is the most challenging goal. Nevertheless, this is a starting point for further studies as some of the main issues are already established. Further investigation of subjects' facial just noticeable difference levels, and the effect from different choice of axes need to be verified. It may

also be interesting to try to model the time it takes to reach the face targets by modelling something similar to Fitts' Law.

The configural approach, though not a popular method for face retrieval, may offer something that current component retrieval systems lack. At this point, there has not yet been an investigation that compares configural and component methodologies for face navigation to determine under what circumstances which is best. We suspect that hybrid approaches for face navigation, as is typical in colour selection interfaces, will ultimately be the most useful. Our methods can also extend to other multi-dimensional representations where sliders alone are not sufficient.

In this research, we offer a way to retrieve a face by borrowing the colour navigating interface, speculating the usage of the colour metaphor, and making some significant findings related to configural navigation interfaces. Our interface, although built with FaceGen faces, is actually independent from the face generator. Thus, we are optimistic that the methods and results may apply as long as the face generator can show face gradients.

As face generation technology improves and applications for creating new faces in entertainment, computer animation and games become more important, having an easy-to-navigate face space is critical. Our research also has potential applications in complementing existing face retrieval systems for crime solving and can extend to the visualization of higher dimensions. Our ultimate direction though is to create our own installation, called *1,001,001 Faces* (the number 1,001,001 is selected for artistic reasons), to provide a modern version

of *Sanjusangen-do* where all people can find the faces of their lost relatives.

Bibliography

[1] T. R. Alley and M. R. Cunningham. Averaged faces are attractive, but very attractive faces are not average. 2(2):123–125, 1991.

[2] D. Archer (Host). The human face: Emotions, identities and masks. Video cassette. Berkeley, CA, 1996. Extension Center for Media and Independent Learning, University of California.
<http://zzyx.ucsc.edu/~archer/vid1.html>.

[3] P. Ashton (Producer). The real Eve. Video cassette. Bethesda, MD, August 2002. Discovery Channel. Jan 28, 2003.
<http://dsc.discovery.com/convergence/realeve/realeve.html>.

[4] E. Baker and M. Seltzer. The mug-shot search problem, 1997. Jan 18, 2003.
<http://citeseer.nj.nec.com/article/baker98mugshot.html>.

[5] Ellen Baker. *The Mug-Shot Search Problem - A Study of the Eigenface Metric, Search Strategies, and Interfaces in a System for Searching Facial Image Data*. PhD thesis, The Division of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, January 1999.

[6] A. Beatty. Singular Inversions - from single photograph to 3D face model, 2003. May 14, 2003.

<http://www.facegen.com>.

[7] I. Biederman and P. Kalocsai. Neurocomputational bases of object and face recognition. 352:1230–1219, 1997.

[8] V. Blanz and T. Vetter. A morphable model for the synthesis of 3D faces. In *Proceedings of SIGGRAPH'99*, pages 187–194, 1999.

[9] B. Bower. Baby facial. *Science News*, 161(20):307, May 2002.

[10] R. Brunelli and O. Mich. Spotit! and interactive identikit system. In *Proceedings of Graphical Models and Image Processing*, volume 58, pages 399–404, 1996.

[11] A. M. Burton and J. R. Vokey. The face-space typicality paradox: Understanding the face-space metaphor. *Quarterly Journal of Experimental Psychology Section A - Human Experimental Psychology*, 51(3):475–483, 1998.

[12] T. A. Busey. Physical and psychological representations of faces: Evidence from morphing. *Psychological Science*, 9:476–483, 1998.

[13] R. Chellappa, C. L. Wilson, and S. Sirohey. Human and machine recognition: A survey. In *Proceedings of the IEEE*, volume 83, pages 705–740, 1995.

[14] J. Cleese (Host). The human face with John Cleese. Video cassette. London, UK, 2001. BBC. Jan 28, 2003. <http://www.bbc.co.uk/science/humanbody/humanface/index. shtml>.

[15] B. H. Cohen. *Explaining Psychological Statistics*. John Wiley & Sons, New York, 2001.

[16] J. O. Cole. *About Face*. MIT Press, Cambridge, Massachusetts, 1998.

[17] D. DeCarlo, D. Metaxas, and M. Stone. An anthropometric face model using variational techniques. In *Proceedings of SIGGRAPH'98*, pages 67–74, 1998.

[18] S. DiPaola. Investigating face space. In *SIGGRAPH'02 Conference Abstracts and Applications*, page 207, 2002.

[19] C. S. Dodson, M. K. Johnson, and J. W. Schooler. The verbal overshadowing effect: Why descriptions impair face recognition. *Memory and Cognition*, 25(2):129–139, 1997.

[20] P. Ekman. *What the face reveals: basic and applied studies of spontaneous expression using the facial action coding system (FACS)*. Oxford University Press, New York, NY, 1997.

[21] K. R. Fehrman and C. Ferhrman. *Color: The Secret Influence*. Prentice Hall, New Jersey, 2000.

[22] S. Fels. Intimacy and embodiment: Implications for art and technology. In *Proceeding of the ACM Workshopson Multimedia*, pages 13–16, Los Angeles, CA, 2000.

[23] P. J. B. Hancock, A. M. Burton, and V. Bruce. Face processing: Human perception and principal components analysis. *Memory & Cognition*, 24:26–40, 1996.

[24] P. J. B. Hancock and C. D. Frowd. Evolutionary generation of faces. In *Proceedings of AISB*, pages 93–99, 1999.

[25] M. Harrison and M. McLennan. *Effective Tcl/Tk Programming*. Addison-Wesley, Massachusetts, 1998.

[26] D. Hopkins. The design and implementation of pie menus. *Dr. Dobb's Journal*, 16(12):16–26, 1991.

[27] A. Hurlbert. Trading faces. *Nature Neuroscience*, 4(1):3–5, 2001.

[28] S. Kawato and J. Ohya. Automatic skin-color distribution extraction for face detection and tracking. In *Proceedings of the 5th International Conference on Signal Processing*, volume II, pages 1415–1418, 2000.

[29] M. M. Klawe. Computer games, education and interfaces: The e-gems project. In *Proceedings of Graphics Interface*, pages 36–39, Kingston, ON, 1999.

[30] A. Klim, F. R. Volkmar, and S. S. Sparrow. *Asperger syndrome*. Guilford Publications Ltd., New York, NY, 2000.

[31] K. R. Laughery and R. H. Fowler. Sketch artist and identi-kit. procedure for recalling faces. *Journal of Applied Psychology*, 65(3):307–316, 1980.

[32] D. A. Leopold, A. J. O'Toole, T. Vetter, and V. Blanz. Prototype-referenced shape encoding revealed by high-level after effects. *Nature Neuroscience*, 4(1):89–94, 2001.

[33] H. Levkowitz. *Color Theory and Modeling for Comupter Graphics, Visualization, and Multimedia applications*. Kluwer Academic Publishers, Boston, 1997.

[34] M. B. Lewis and R. A. Johnston. A unified account of the effects of caricaturing faces. *Visual Cognition*, 6(1):1–41.

[35] E. F. Loftus. *Eyewitness Testimony*. Harvard University Press, Cambridge, Massachusetts, 1944.

[36] I. S. MacKenzie and W. Buxton. Extending Fitts' law to two dimensional tasks. In *Proceedings of the CHI'92 conference on Human Factors in Computing Systems*, pages 219–226, 1992.

[37] J. Marks, B. Andalman, P. A. Beardsley, W. Freeman, S. Gibson, J. Hodgins, T. Kang, B. Mirtich, H. Pfister, W. Ruml, K. Ryall, J. Seims, and S. Shieber. Design galleries: A general approach to setting parameters for computer graphics and animation. In *Proceedings of SIGGRAPH'97*, pages 389–400, 1997.

[38] S. Marquardt. Marquardt beauty analysis, 2001. Nov 30, 2001. <http://www.beautyanalysis.com>.

[39] D. W. Martin. *Doing Psychology Experiments*. Books/Cole Publishing Company, Monterey, California, 1985.

[40] D. McNeill. *The Face*. Little, Brown and Company, Boston, 1998.

[41] A. N. Meltzoff and M. K. Moore. Imitation in newborn infants: Exploring the range of gestures imitated and the underlying mmechanisms. *Developmental Psychology*, 25(6):954–962, 1989.

[42] B. F. Miller and C. B. Keane. *Encyclopedia and dictionary of medicine and nursing*, page 1223. W. B. Saunders Company, Philadelphia, 1992.

[43] D. C. Montgomery. *Design and Analysis of Experiments*. John Wiley & Sons, New York, 2001.

[44] C. Nass, E. Y. Kim, and E. J. Lee. When my face is the interface: An experimental comparison of interacting with one's own face or someone else's face. In *Proceedings of SIGCHI*, volume 1, pages 148–154, Los Angeles, CA, April 1998.

[45] C. Newman. The enigma of beauty. *National Geographic*, 197(1):94–121, January 2000.

[46] A. J. O'Toole, H. Abdi, K. A. Deffenbacher, and D. Valentin. Low-dimensional representation of faces in higher dimensions of the face space.

Journal of the Optical Society of America A - Optics Image Science and Vision, 10(3):405–411, 1993.

[47] P. S. Penev and L. Sirovich. The global dimensionality of face space. In *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*, pages 264–270, 2000.

[48] A. Pentland, R. W. Picard, and S. Sclaroff. Photobook: Tools for content-based manipulation of image databases. In *Proceedings SPIE Storage and Retrieval Image and Video Databases II*, volume 2185, 1994.

[49] R. Plutchik. *The Emotions: Facts, Theories, and a New Model*. Random House, New York, 1962.

[50] M. J. Roberts and R. Russo. *A Student's Guide to Analysis of Variance*. Routledge, New York, 1999.

[51] J. W. Schooler and T. Y. Engstler-Schooler. Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology*, 22(1):36–71, 1990.

[52] M. W. Schwarz, W. B. Cowan, and J. C. Beatty. An experimental comparison of RGB, YIQ, LAB, HSV and opponent color models. *ACM Transactions on Graphics*, 6(2):123–158, 1987.

[53] B. Shneiderman. Direct manipulation: A step beyond programming languages. In *IEEE Computer*, volume 16, pages 57–69, 1983.

[54] J. L. Swerdlow. Unmasking skin. *National Geographic*, 202(5):36–63, November 2002.

[55] M. A. Tapia and G. Kurtenbach. Some design refinements and principles on the appearance and behavior of marking menus. In *Proceedings of UIST*, pages 189–195, 1995.

[56] Bernard Tiddeman and David Perrett. Moving facial image transformations based on static 2D prototypes. In V. Skala, editor, *WSCG 2001 Conference Proceedings*, 2001.

[57] C. Tredoux, Y. Rosenthal, L. da Costa, and D. Nuenz. Face reconstruction using a configural, eigenface-based composite system. In *Proceedings of 3rd Biennial Meeting of Society of Applied Research in Memory and Cognition*, Boulder, Colorado, 1999.

[58] A. Treves. On the perceptual structure of face space. *Biosystems*, 40(1-2):189–196, 1997.

[59] M. Turk and A. Pentland. Face recognition using eigenfaces. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 586–591, 1991.

[60] T. Valentine. A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Quarterly Journal of Experimental Psychology A*, 43:161–204, 1991.

[61] T. Valentine. Face-space models of face recognition. In M. J. Wenger and J. T. Townsend, editors, *Computational, geometric, and process perspectives on facial recognition: Contexts and challenges*, pages 83–113. Lawrence Erlbaum Associates, Mahwah, New Jersey, 2001.

[62] J. H. Walker, L. Sproull, and R. Subramani. In *Proceedings of SIGCHI*, pages 85–91, 1994.

[63] W. Weiten. *Psychology: themes and variations*. Pacific Grove, California, 1992.

[64] S. Wells. *The Journey of Man: a Genetic Odyssey*. Penguin Press, London, England, 2002.

[65] M. Woo, J. Neider, T. Davis, and D. Shreiner. *OpenGL Programming Guide*. Addison Wesley Longman, Massachusetts, third edition, 1999.

[66] J. K. Wu, Y. H. Ang, P. Lam, H. H. Loh, and A. Desai Narasimhalu. Inference and retrieval of facial images. *Multimedia Systems*, 2(1):1–14, 1994.

[67] A. W. Yip and P. Sinha. Contribution of color to face recognition. *Perception*, 31(5):995–1003, 2002.

[68] A. W. Young. *Face and Mind*. Oxford University Press, Oxford, 1998.

Appendix A

Interview Questions

A.1   Pilot 1

1. Which of the two interfaces (wheel and static slider) do you prefer, and why?

2. Do you find it harder to navigate with six axes compared to three?

3. Do you have any additional comments?

A.2   Pilot 2

1. Which of the two interfaces (wheel and dynamic slider) do you prefer, and why?

2. Do you notice a difference between the correlated and uncorrelated axes?

3. Do you have any additional comments?

A.3   Pilot 3 and the Final Experiment

1. Which of the two interfaces (wheel and dynamic slider) do you prefer, and why?

2. Do you find it harder to navigate under a finer resolution condition? If so, what would you do to improve it?

3. Do you have any additional comments?

Appendix B

Analysis of Variance

In all our experiments, we use within-subject factorial designs. Factorial design calls for application of ANOVA. We have followed within-subject ANOVA formulas in [43], [50] and *HyperStat*[1] with six dependent measurements (where applicable) to find out which variables are significant. The dependent measurements we use are time[2], accuracy score[3], repetition[4], uniqueness[5], refinement efficiency[6] and approaching efficiency[7]. Although ANOVA is a useful statistic technique, we use it mainly to sift out significant variables—particularly the interface factor. We include the ANOVA test results here. For a statistical report and interpretations, readers can see chapters 6 to 9.

Appendix B includes only those ANOVA tables where significant variables are detected. Those that do not are excluded. "Simple main effect analysis" tables are also included for the variables that show significant interaction. Sig-

---

[1] HyperStat is at http://davidmlane.com/hyperstat/index.html. (Last accessed April 21, 2003.)

[2] Time is measured in seconds per trial.

[3] Score is assigned in such a way that if the final target takes, at most, n number of steps and subjects only manage to come to its immediate neighbourhood, i.e. one step away from the target, then score is $\frac{n-1}{n}\%$. Similarly, if subjects arrive at k step away from the target, the score is $\frac{n-k}{n}\%$. Note that step is the step in the discretized face space we devise for user testing, not the step size of *FaceGen* face space.

[4] The repetition measure is determined by the number of face locations subjects revisit.

[5] Uniqueness is measured by the number of unique faces subjects encounter per trial.

[6] Refinement efficiency is measured by the number of face locations visited that fall within two steps from the face target.

[7] Approaching efficiency is measured in terms of the number of face locations visited when subjects are more than two steps away from the face targets.

	Time	Score	Repetition
Pilot 1	none	none	interface
Pilot 2	none	none	none
Pilot 3	resolution	none	resolution
Final Experiment	resolution target	resolution	resolution interface interface × resolution

	Unique Face counts	Face counts while refining	Face counts while approaching
Pilot 1	interface	N/A	N/A
Pilot 2	N/A	none	N/A
Pilot 3	N/A	resolution interface × resolution	resolution interface target interface × resolution
Final Experiment	N/A	resolution	resolution interface target

Table B.1: The table lists the variables found to be significant from within-subject ANOVA. "N/A" means that the particular dependent measurement is not used in ANOVA testing. "None" means no variables are found to be significant by ANOVA. Of interest is where interface is found to be significant because it indicates differences between the interfaces when performance is measured in some measurements under particular conditions.

nificant variables detected by ANOVA are summarized in Table B.1.

These tests are also used to accept or reject the hypotheses tested on the interface variable. They are presented in Chapter 5. Hypotheses testing results are summarized in Table B.2.

B.1 Pilot 1

In our first pilot test, we run ANOVA on subjects' performance measured in time, score, repetition and uniqueness counts. The ANOVA results allow us to see which variables cause significant variance and the interface variable is of most interest.

	H_{01} (Time)	H_{02} (Score)	H_{03} (Repetition)
Pilot 1	accept	accept	reject
Pilot 2	accept	accept	accept
Pilot 3	accept	accept	accept
Final Experiment	accept	accept	reject

	H_{04} (Unique face counts)	H_{05} (Visits while refining)	H_{06} (Visits while approaching)
Pilot 1	reject	N/A	N/A
Pilot 2	N/A	accept	N/A
Pilot 3	N/A	accept	reject
Final Experiment	N/A	accept	reject

Table B.2: Summary of hypothesis testing. "N/A" means that the hypothesis is not tested because it is inapplicable. "Accept" implies we do not find significance in rejecting the null hypotheses as the p-values for the interface variable are not small enough. "Reject" means we do find sufficient significance in rejecting the null hypotheses.

Source Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P-Value
Subject	235.83	5	47.17		
Interface	816.67	1	816.67	11.29	0.0037
Error	1229.33	17	72.314		
Total	2281.83	23			

Table B.3: One-way ANOVA of the number of revisits of faces in the face space across all trials in Experiment 1.

Only by using the measurements of the subjects' performances in repetitiveness and uniqueness is the interface variable significant. Their ANOVA tables are listed in Table B.3 and B.4 respectively. We also run an extra ANOVA, shown in Table B.5, using the amount of time people spend looking at every face. Again, interface is a significant factor.

Source Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P-Value
Subject	6138.8	5	1227.77		
Interface	46464.00	1	46464.00	38.42	0
Error	20558.50	17	1209.32		
Total	73161.30	23			

Table B.4: One-way ANOVA of the number of unique faces subjects saw across all trials in Experiment 1.

Source Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P-Value
Subject	22.48	5	4.50		
Time/Face	91.70	1	91.70	85.35	0
Error	18.265	17	1.07		
Total	132.44	23			

Table B.5: One-way ANOVA of the time spent on every unique face across all trials of Experiment 1.

B.2 Pilot 2

In Pilot 2 ANOVA tests detect no significance of any variables with any dependent measurements.

B.3 Pilot 3

In Pilot 3, we use ANOVA on subjects' performances measured in time, score, repetitiveness, refinement efficiency and approach efficiency. The ANOVA done on score measurements is not included because it shows no significance of any variables. The interface factor is not significant in ANOVA Tables B.6 and B.7. However, it is significant in Table B.10, and its interaction with resolution is significant in Tables B.8 and B.10. Their follow up "simple main effect analysis" results are shown in Tables B.9 and B.11, respectively, which further scrutinize

Source Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P-Value
Subject	112615.30	5	22523.10		
Interface	8782.20	1	8782.20	0.70	0.44
Error	62397.80	5	12479.60		
Resolution	500398.00	1	500398.00	46.28	0.001
Error	54066.40	5	10813.30		
Target	30263.00	1	30263.00	5.24	0.07
Error	28900.70	5	5780.10		
Interface × Resolution	0.50	1	0.50	0.00	0.99
Error	45324.10	5	9064.80		
Interface × Target	18338.60	1	18338.60	1.86	0.23
Error	49360.70	5	9872.10		
Resolution × Target	14347.30	1	14347.30	1.05	0.35
Error	68625.90	5	13725.2		
Interface × Resolution × Target	5501.30	1	5501.30	0.92	0.34
Error	315605.3	53	5954.8		
Total	1314527.1	95			

Table B.6: Three-way ANOVA of the time taken across all trials in Experiment 3.

the interaction between interface and resolution variables.

Note that we do not use the unique face counts in Pilot 3 as a dependent measurement because the wheel and the dynamic slider interface work similarly in providing users with gradients of faces (the wheel interface displays one step neighbours, and the dynamic slider interface displays the faces at extreme values), and the unique face counts are not significantly different.

Source Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P-Value
Subject	52.46	5	10.49		
Interface	0.04	1	0.04	0.01	0.94
Error	29.46	5	5.90		
Resolution	135.38	1	135.38	11.4	0.02
Error	59.38	5	11.88		
Target	1.04	1	1.04	0.10	0.76
Error	50.46	5	10.09		
Interface × Resolution	1.04	1	1.04	0.57	0.49
Error	9.21	5	1.84		
Interface × Target	1.04	1	1.04	0.35	0.58
Error	14.96	5	2.99		
Resolution × Target	26.04	1	26.04	2.08	0.21
Error	62.71	5	12.54		
Interface × Resolution × Target	1.04	1	1.04	0.24	0.63
Error	230.71	53	4.35		
Total	674.96	95			

Table B.7: Three-way ANOVA of the number of revisits subjects made in the face space for Experiment 3.

Source Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P-Value
Subject	94.93	5	18.99		
Interface	86.26	1	86.26	2.20	0.20
Error	196.43	5	39.29		
Resolution	380.01	1	380.01	19.11	0.01
Error	99.43	5	19.89		
Target	0.01	1	0.01	0	0.98
Error	55.68	5	11.14		
Interface × Resolution	114.84	1	114.84	10.92	0.02
Error	52.59	5	10.52		
Interface × Target	3.76	1	3.76	0.17	0.70
Error	111.43	5	22.29		
Resolution × Target	55.51	1	55.51	1.91	0.23
Error	145.43	5	29.09		
Interface × Resolution × Target	3.76	1	3.76	0.17	0.68
Error	1178.68	53	22.24		
Total	2578.74	95			

Table B.8: Three-way ANOVA of the faces visited at the refining phase in Experiment 3.

Source Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P-Value
Interface at					
Coarse Resolution	0.51	1	0.51	0.03	0.88
Fine Resolution	100.04	1	100.04	5.09	0.07
Error	98.21	5	19.64		
Resolution at					
Wheel	19.26	1	19.26	1.94	0.22
Dynamic Slider	228.17	1	228.17	22.95	0.005
Error	49.71	5	9.94		

Table B.9: Simple main effect analysis of face locations visited during refinement phase in Pilot 3.

Source Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P-Value
Subject	243.18	5	48.64		
Interface	585.09	1	585.09	7.84	0.04
Error	373.34	5	74.67		
Resolution	1312.76	1	1312.76	36.63	0.00
Error	179.18	5	35.84		
Target	326.34	1	326.34	12.74	0.02
Error	128.09	5	25.62		
Interface × Resolution	283.59	1	283.59	7.73	0.04
Error	183.34	5	36.67		
Interface × Target	86.26	1	86.26	2.34	0.19
Error	184.68	5	36.94		
Resolution × Target	7.59	1	7.59	0.09	0.78
Error	438.34	5	87.67		
Interface × Resolution × Target	31.51	1	31.51	1.45	0.23
Error	1151.43	53	21.73		
Total	5514.74	95			

Table B.10: Three-way ANOVA of the faces visited at the approaching phase in Experiment 3.

Source Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P-Value
Interface at					
Coarse Resolution	13.5	1	13.5	0.36	0.57
Fine Resolution	420.84	1	420.84	11.27	0.02
Error	186.67	5	37.33		
Resolution at					
Wheel	704.17	1	704.17	39.30	0.002
Dynamic Slider	94.01	1	94.01	5.25	0.07
Error	89.59	5	17.92		

Table B.11: Simple main effect analysis of face locations visited during approaching phase in Pilot 3.

Source Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P-Value
Subject	461087.10	14	32934.80		
Interface	1133.30	1	1133.30	0.21	0.66
Error	77397.20	14	5528.40		
Resolution	1260838.90	1	1260838.90	88.86	0.00
Error	198656.80	14	14189.80		
Target	124732.40	1	124732.40	8.21	0.01
Error	212658.50	14	15189.90		
Interface × Resolution	15300.90	1	15300.90	1.52	0.24
Error	141001.70	14	10071.60		
Interface × Target	5662.70	1	5662.70	0.98	0.34
Error	80562.60	14	5754.50		
Resolution × Target	9178.30	1	9178.30	0.66	0.43
Error	194645.60	14	13903.30		
Interface × Resolution × Target	7723.90	1	7723.90	0.92	0.34
Error	1122976.30	134	8380.40		
Total	3913556.10	239			

Table B.12: Three-way ANOVA of the time taken in the final experiment.

B.4 Final Experiment

In our final experiment, we use the same test design as in Pilot 3. We run ANOVA on subjects' performances measured in time, score, repetitiveness, refinement efficiency and approach efficiency. Interface is not a significant factor in ANOVA Table B.12, B.13 and B.16. However, the interface factor is a significant source of variation when the dependent measurements are repetition and approach efficiency counts—as indicated by Tables B.14 and B.17 respectively. To further investigate interaction between resolution and interface from Table B.14, a "simple main effect analysis" is shown in Table B.15.

Source Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P-Value
Subject	1.81	14	0.13		
Interface	0.00	1	0.00	0.00	0.99
Error	0.70	14	0.05		
Resolution	1.30	1	1.3	16.87	0.001
Error	1.08	14	0.08		
Target	0.01	1	0.01	0.23	0.64
Error	0.47	14	0.03		
Interface × Resolution	0.09	1	0.09	1.79	0.20
Error	0.69	14	0.05		
Interface × Target	0.19	1	0.19	3.19	0.10
Error	0.82	14	0.06		
Resolution × Target	0.00	1	0.00	0.08	0.79
Error	0.67	14	0.05		
Interface × Resolution × Target	0.04	1	0.04	1.62	0.21
Error	3.64	134	0.03		
Total	11.48	239			

Table B.13: Three-way ANOVA of the score in the final experiment.

Source Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P-Value
Subject	858.61	14	61.33		
Interface	136.50	1	136.50	5.28	0.04
Error	361.81	14	25.84		
Resolution	670.00	1	670.00	28.72	0.0001
Error	326.56	14	23.33		
Target	15.50	1	15.50	0.8	0.39
Error	271.81	14	19.42		
Interface × Resolution	130.54	1	130.54	7.08	0.02
Error	258.03	14	18.43		
Interface × Target	22.20	1	22.20	1.50	0.24
Error	207.61	14	14.83		
Resolution × Target	2.60	1	2.60	0.12	0.73
Error	293.46	14	20.96		
Interface × Resolution × Target	0.94	1	0.94	0.04	0.84
Error	3066.63	134	22.89		
Total	6622.80	239			

Table B.14: Three-way ANOVA of the revisits in the final experiment.

Source Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P-Value
Interface at					
Coarse Resolution	0.02	1	0.02	0.00	0.98
Fine Resolution	133.50	1	133.50	3.69	0.11
Error	180.90	5	36.18		
Resolution at					
Wheel	348.00	1	348.00	13.36	0.01
Dynamic Slider	52.27	1	52.27	2.01	0.22
Error	130.28	5	260.07		

Table B.15: Simple main effect analysis of revisits in the final experiment.

Source Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P-Value
Subject	1406.15	14	100.44		
Interface	45.94	1	45.94	1.95	0.18
Error	329.00	14	23.50		
Resolution	862.60	1	862.60	34.18	0.00
Error	353.33	14	25.24		
Target	33.00	1	33.00	0.31	0.59
Error	1488.18	14	106.30		
Interface × Resolution	21.00	1	21.00	0.39	0.54
Error	746.18	14	53.3		
Interface × Target	42.50	1	42.50	1.73	0.21
Error	343.43	14	24.53		
Resolution × Target	45.94	1	45.94	1.11	0.31
Error	579.00	14	41.36		
Interface × Resolution × Target	130.54	1	130.54	2.26	0.14
Error	7743.15	134	57.79		
Total	14169.96	239			

Table B.16: Three-way ANOVA of the faces visited during the refinement stage in the final experiment.

Source Variation	Sum of Squares	Degrees of Freedom	Mean Square	F_0	P-Value
Subject	2689.00	14	192.07		
Interface	908.70	1	908.70	16.81	0.00
Error	756.70	14	54.05		
Resolution	3896.20	1	3896.20	34.96	0.00
Error	1572.20	14	112.30		
Target	1535.20	1	1535.20	24.04	0.00
Error	894.00	14	63.86		
Interface × Resolution	80.50	1	80.50	2.77	0.12
Error	407.20	14	29.08		
Interface × Target	133.50	1	133.50	2.69	0.12
Error	694.90	14	49.64		
Resolution × Target	343.20	1	343.20	3.39	0.09
Error	1418.20	14	101.30		
Interface × Resolution × Target	90.00	1	90.04	1.69	0.20
Error	7124.1	134	53.17		
Total	22543.80	239			

Table B.17: Three-way ANOVA of the faces visited during the approaching stage in the final experiment.

Appendix C

Additional Graphs of Navigation Pattern Analysis

By analyzing the number of times subjects arrive within a certain range from the target face in Pilot 3, we find that the sliders tend to exceed the wheel interface in the number of face locations visited within the range of $[0, 2]$ steps. Although the curves appear irregular, this tendency persists in the final experiment. Figures C.1, C.2, C.3 and C.4 show the trend in the following four conditions: coarse resolution with a target, fine resolution with a near target, coarse resolution with a far target and fine resolution with a far target.
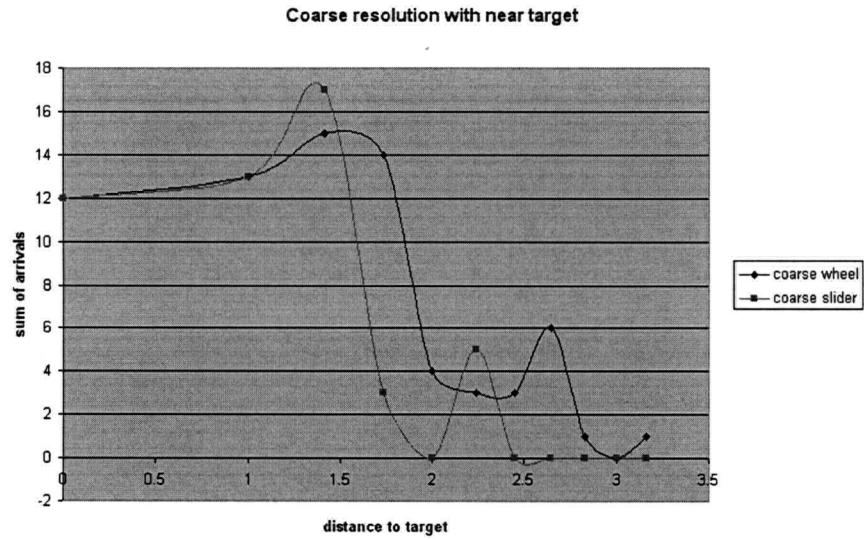
Coarse resolution with near target

Figure C.1: Navigation arrival pattern for coarse resolution with a near target condition in Pilot 3.
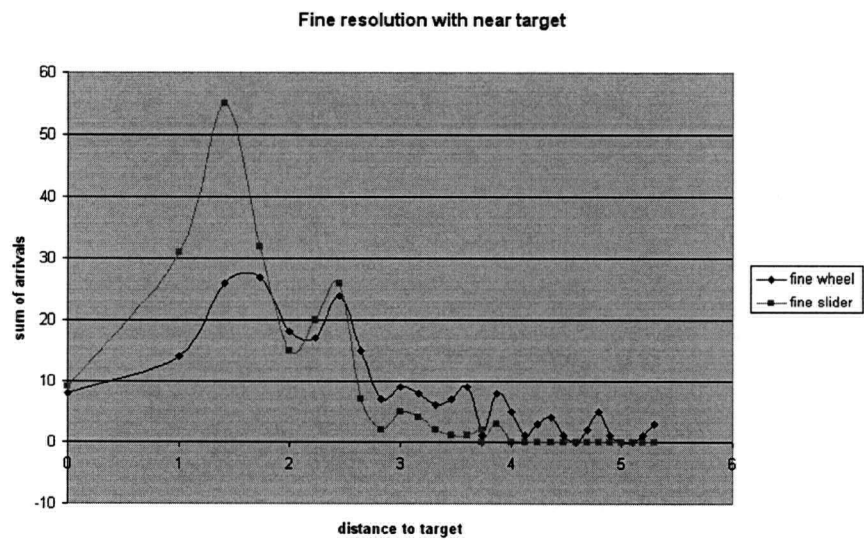
Fine resolution with near target

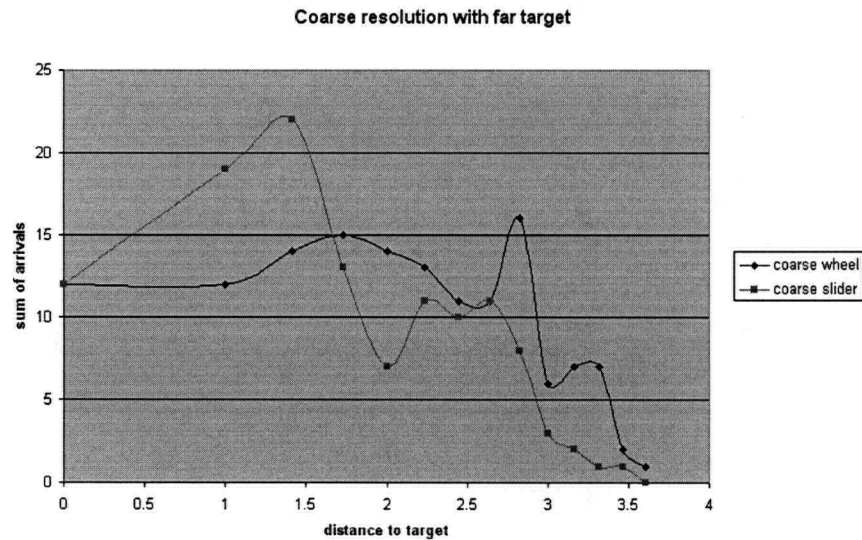Figure C.2: Navigation arrival pattern for fine resolution with a near target condition in Pilot 3.

Coarse resolution with far target

Figure C.3: Navigation arrival pattern for coarse resolution with a far target condition in Pilot 3.
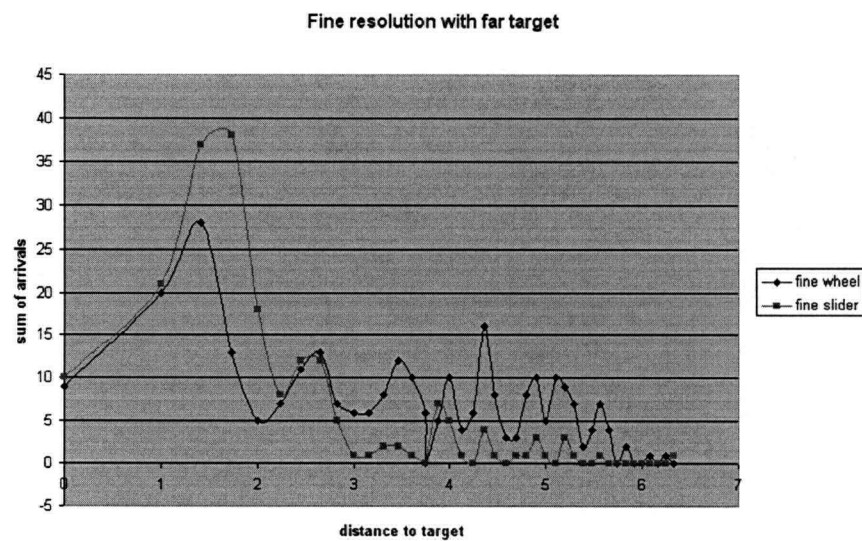
Fine resolution with far target

Figure C.4: Navigation arrival pattern for fine resolution with a far target condition in Pilot 3.