Depth in Dedicated Mobile Device User Interfaces for Auto-Stereoscopic Displays

Master Thesis

Sin Lin Wu
16 February 2010

TU Delft

MOTOROLA

Depth in Dedicated Mobile Device User Interfaces for Auto-Stereoscopic Displays

THESIS

submitted in partial fulfillment of the

requirements for the degree of

MASTER OF SCIENCE

in

MEDIA AND KNOWLEDGE ENGINEERING

by

Sin Lin Wu

Born in Amsterdam, the Netherlands

16 February 2010

Man-Machine Interaction Group
Faculty of Electrical Engineering, Mathematics
and Computer Science
Delft University of Technology
The Netherlands

Motorola, Inc
Tempe, Arizona
United States of America

Man-Machine Interaction Group
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology
Mekelweg 4
2628 CD Delft
The Netherlands

Motorola, Inc
2900 South Diablo Way
Tempe, AZ 85282
United States of America

Graduation Committee

Prof. dr. Ingrid E.J.R. Heynderickx	TU Delft, Faculty of EEMCS, MMI group
Dr. ir. Willem-Paul Brinkman	TU Delft, Faculty of EEMCS, MMI group
Dr. Sylvia C. Pont	TU Delft, Faculty of Industrial Design, HICD

Preface

This thesis describes the research that has been conducted to accomplish my master project. This project was carried out at three places. One part was carried out at the Man-Machine Interaction Group, of the faculty of Electrical Engineering, Mathematics and Computer Science, University of Technology Delft. The other parts were carried out at Motorola, Inc and Arizona State University in Tempe, Arizona.

First of all, I would like to thank from the bottom of my heart the most important person during my master project, my thesis supervisor, prof. dr. Ingrid Heynderickx, for her guidance and support. I would also like to thank her for introducing me to the subject of auto-stereoscopic displays and my master thesis subject.

From Motorola, Inc I would like to thank Bernard Coll for making it possible for me to visit the Motorola lab and for his support and guidance. I would also like to thank Michael Johnson for everything he has done for me while I was in Tempe. I would like to thank all the other people at Motorola for making me feel comfortable at the lab.

I would like to thank dr. Lina Karam of the Department of Electrical Engineering at Arizona State University for asking her students to participate in my experiments.

To be able to do this project a number of experiments had to be done. I would like to thank all the participants of my experiments for taking their time to participate.

Finally, I would like to thank my family and friends who supported me during the whole duration of this project.

Sin Lin Wu

Delft, February 2010

Summary

With the growing usage of mobile devices, manufacturers have to distinguish themselves in their devices to keep their share of the market. Most important is that the consumer is satisfied with a given brand, since then the chances are high that he will stay loyal to that brand when it is time for a new device. A consumer buys a mobile device based on several factors. The user interface is one of these factors, and it is an easy aspect for the consumer to evaluate. It is what he sees and uses, and therefore, has to be made easy to use, intuitive and attractive. One way of making a user interface attractive is by using a new technology, such as an auto-stereoscopic display. This type of display makes use of binocular depth, also called stereopsis, which allows users to see depth without the necessity of using viewing aids.

Nowadays, static images on a standard 2D display get an impression of depth by using monocular depth cues, e.g. occlusion, shadows, relative size, and linear perspective. However, an auto-stereoscopic display provides a real sense of depth with objects popping out of the display screen. It is interesting to investigate the effect on the overall experience of a user interface of a mobile device with an auto-stereoscopic display when combining stereopsis with these monocular depth cues. To answer this research question three experiments have been performed.

The goal of the first experiment – the pilot experiment – was to find which attributes were appropriate for evaluating the overall experience of a graphical user interface of a mobile device. In this experiment subjects saw stimuli based on a simplified menu of a mobile device, consisting of a background image with five background icons, and one 'selected icon' on the foreground. The stimuli varied in the monocular depth cue used, being shadowing, relative size, and luminance differences, and in the background image and depth mode used. The subjects were asked to score these stimuli on three predefined attributes, which were affordance, aesthetics, and preference. From this experiment the attributes affordance and aesthetics were found to represent different assessment criteria, which implied that these two attributes were useful for the assessment of a graphical user interface of a mobile device.

The goal of the second experiment – the tuning experiment – was to find the preferred setting of the subjects for the disparity difference between foreground and background items in the graphical user interface, for the size increase of the 'selected icon' and for the luminance increase of the 'selected icon'. Subjects were requested to indicate their preferred setting via a tuning for each of the variables separately. The results showed that there was quite some spread in preferred disparity among the subjects. The preferred size of the 'selected icon' was on average about 130% of its original size. The increase in luminance for the 'selected icon' was found to be too small to be detected in the absence of reference material. Therefore, it was decided to leave this variable out of the rest of the study.

The goal of the third experiment, being the main experiment, was to find the added value of the various depth cues for the user interface of a mobile device with an auto-stereoscopic display. In this experiment the subjects were selected based on having about the same preferred disparity. These subjects were asked to score the stimuli on the attributes perceived amount of depth, image quality, affordance, and aesthetics. The stimuli consisted of combinations of the same variables as in the first experiment, with the exception of the luminance cue. The results showed that the bigger size of the 'selected icon' and the addition of a shadow both contributed to higher scores for all four attributes. Additionally, the depth mode 'sinking background', in which the 'selected icon' was displayed at the screen and the background behind the screen, gained the highest scores for the perceived amount of

depth, image quality and affordance. The depth mode 'none', representing the background image that already contained depth had the highest score on aesthetics. The natural background image had the highest score on perceived depth and affordance, while the uniform background had the highest score on image quality.

Overall it can be concluded that an auto-stereoscopic display has clearly added value for the design of a graphical user interface on a mobile device. The various variables studied for the graphical user interface affected different aspects of its overall experience. Clearly, a 'sinking background' depth mode is more appreciated than a 'floating icon' depth mode. A background containing depth is clearly appreciated from an aesthetic point of view. Additionally, the user interface should be designed such that no artifacts affecting the overall image quality are introduced.

As the amount of disparity that has to be used is very dependent on the viewer, it may be a good option to allow the user of a mobile device to personalize the amount of disparity to his or her own preference.

Table of contents

1 Introduction

Over the last decade the usage of mobile devices has been rising tremendously. When looking at the number of mobile cellular subscribers per 100 people, the worldwide mobile phone market is still growing (ITU, 2009). Figure 1.1 shows that there were less than 500 million subscribers worldwide before 1999 and this number grew to approximately 4 billion in 2008.

A mobile phone is easy to carry because of its size and weight. Therefore, mobile phones are accessible at any time. The technology of the mobile phone has been evolving in a way that it is possible to do more than only call with a mobile phone. People can use the mobile phone to access information on the internet, make a phone call while travelling or just send a quick text message.

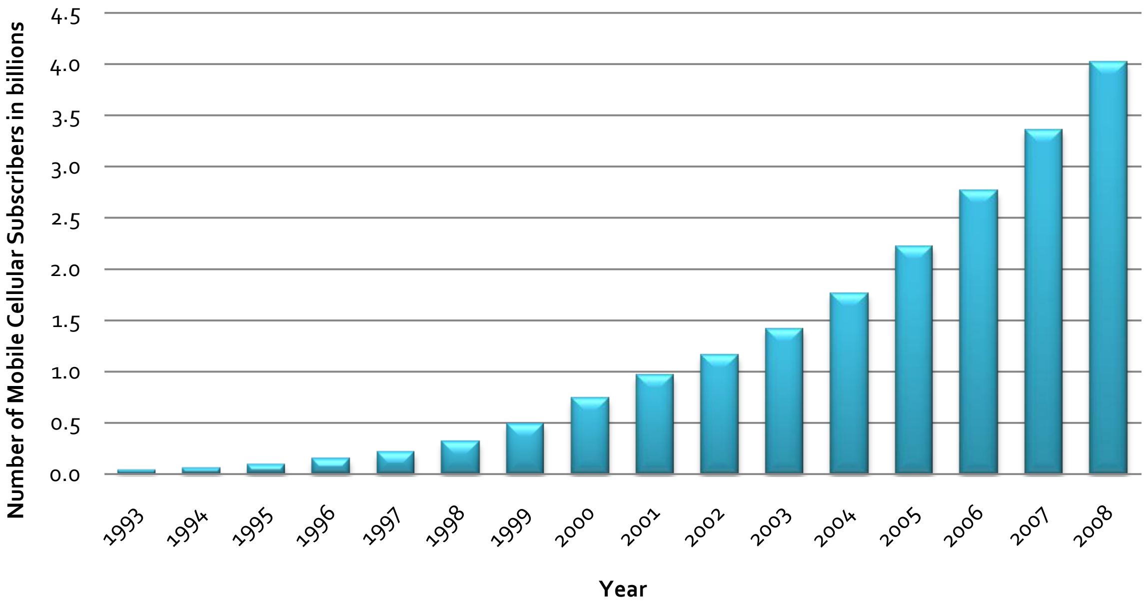


Figure 1.1 Worldwide trend of mobile cellular subscribers per 100 people according to the statistics of the International Telecommunication Union

Because of this growing market, there are now many mobile phone manufacturers, the competition amongst which is high. Manufacturers have to distinguish themselves in their phones to keep their share of the market. The basic requirements of a mobile phone on performance and functionality are met by most of the devices; therefore, consumers take other factors in consideration. The decision of a consumer to buy a certain brand depends on factors such as price, functionality, brand name, perceived ease-of-use and robustness, exposure in shops and advertising, ratings and recommendations, and finally attractiveness of the design of the device (Sparre, 2007).

The user interface is an easy aspect of the mobile device to evaluate for the user. It is what they see and what they use. Therefore, the design of the user interface is an important aspect. It has to be easy to use, intuitive and attractive.

1.1 The user interface

The user interface includes those parts of hardware with which the user can interact with a device. For a computer, e.g. the user interface includes the screen, keyboard, mouse, and the images on the screen, such as the menus and windows (Redmond-Pyle & Moore, 1995). For a mobile device the user interface comprises the screen as well as the device itself. The menus, windows, and icons that appear on the screen are the graphical components of the graphical user interface. The graphical user interface is a way for humans to interact with the computer, by means of clicking on the graphical components. Other ways of interaction are, for example, the command line interface. Here the user interacts with the computer by typing in commands. The graphical user interface is easier to use, as the function behind an icon is a graphical representation of a physical object, and thus very recognizable. The user can interact with the computer without having to remember the commands. Mobile devices mainly make use of a graphical user interface, and a command line interface is usually not available.

The most critical quality aspect of a user interface is its usability as its purpose is to make a device usable to the user (Redmond-Pyle & Moore, 1995).

When a device has an attractive user interface, a consumer will consider buying that device more strongly than when the device has a less attractive design and user interface (Sparre, 2007). One of the options to make a user interface attractive is personalization. Most brands of mobile devices allow users to personalize the appearance of the user interface to match it to their own needs and desires (Sparre, 2007). This means that users can, among other options, set their own background, choose between a text-based or icon-based interface, and choose a ring tone. Some brands allow users to rearrange the menu icons for faster access.

If a consumer is satisfied with a given brand, the chances are high that the consumer will be loyal to that brand when it is time for a new phone. Manufacturers have to distinguish themselves to be able to compete. This can be done by using new technologies.

1.2 New technologies for mobile devices

When mobile phones were just introduced to the consumer market, they were only used for calling and text messaging. Nowadays, more and more smartphones are available on the market. Smartphones are mobile phones with computer functions, such as browsing on the internet and sending e-mails. With this change in technology, the technology used for user interfaces needs to be adapted too. An evolving technology is the input method of the mobile phone. It has been evolving from a simple mobile phone keypad with 10 keys with digits and two additional keys, to QWERTY keyboards and touch screens.

Another new technology, recently introduced in mobile devices, is auto-stereoscopic displays. As a consequence users can see objects hanging in front or behind their display. This technology is based on the phenomenon of stereopsis. Stereopsis was described by Charles Wheatstone in 1838 as "… the mind perceives an object of three-dimensions by means of the two dissimilar pictures projected by it on the two retinae…" (Wheatstone, 1838). Besides the definition of stereopsis, Charles Wheatstone also described the effects and human factors of stereopsis, and built the first stereoscope. Human beings use stereopsis in their daily life to perceive depth in their environment. Both eyes receive a slightly different perspective of a scene on the retina and the brain fuses both images. The relative depth information is extracted from the retinal disparity, which is the difference in distance on the retina between corresponding points in the two images. Displays with stereopsis have the ability to visualize depth in a true manner, whereas 2D displays create the impression of depth with visual depth cues, such as occlusion, shadow and relative size.

Unlike stereopsis which is a binocular cue, depth cues like occlusion, shadow, and relative size are monocular cues, which refer to the fact that an impression of depth can be obtained based on vision with one eye. The viewer gets a depth impression with occlusion, since he knows that when one object partly hides another object, the object in the front is closer than the hidden object. Shadows give information about the shape of the object and location of the object with respect to the surface in a scene. Relative size gives depth information as the viewer knows that larger objects have to be closer to him than smaller objects. In this study monocular cues will be combined with stereopsis to see if there is any added effect on the perceived depth.

Over the years several technologies have been developed where viewing aids are necessary in order to perceive 3D, e.g. the technology used in cinemas where you have to wear glasses. But recently displays that do not require viewing aids have been introduced onto the market. These displays are called auto-stereoscopic displays. The biggest advantage of this kind of displays is that they do not require viewing aids, and therefore, are more practical when used for example for mobile devices. Users of this kind of devices will not have to carry a pair of glasses around to be able to use their device. In 2002, the first mobile phone with an auto-stereoscopic display, the Sharp mova SH251iS, reached the consumer market in Japan (NTTDocomo, 2002). Later a second phone with an auto-stereoscopic display, the Sharp mova SH505i, appeared on the market, also by NTT DoCoMo (Sharp, 2003). Both phones, however, disappeared again from the market due to the lack of a killer application for the 3D technology.

Nowadays with the increased functionality in mobile phones an auto-stereoscopic display can have added value for e.g. the gaming experience of the user (Rajae-Joordens, 2008). It can also facilitate navigation, by showing the roads with landmarks in 3D. Displaying videos and images can also benefit from an auto-stereoscopic display. The depth in the contents can affect the viewing experience, and more so the naturalness of the content (Seuntiëns, et al., 2005). But using 3D can also improve the design of the menu interfaces. The third dimension can give a more organized feeling, because it is possible to put menu icons on different depth layers. The different depth layers can be used to put for example the most important or selected icon on the topmost layer and the rest on the layers behind this one. For file management on a personal computer a technique called Data Mountain exists. Data Mountain allows users to arrange their documents, in the form of thumbnail images, on an inclined 3D plane (Robertson, et al., 1998). Another 3D document manager is the Task Gallery (Robertson, et al., 2000). In an experiment where different groups of subjects had to group web pages together and later locate them in physically and virtually 2D, $2^1/_2$D and 3D planes, it was found that with the increase of use of the third dimension, the ability of locating a web page deteriorated. The subjects also found the 3D interfaces more cluttered and less efficient (Cockburn & McKenzie, 2002).However, in another experiment it was concluded that there was no significant difference between task performance in 2D and 3D interfaces, but that nonetheless, there was a significant preference for 3D interfaces (Cockburn & McKenzie, 2001). Another study compared 2D and various modes of 3D in task involving the comprehension of 3D graphs. Here the 3D conditions outperformed the 2D condition. The ability to be able to move or rotate the graph was proved to be a valuable feature in 3D (Ware & Franck, 1996). The Space Manager was designed for a mobile terminal (Hakala, Lehikoinen, & Aaltonen, 2005). It is a document manager application that uses a tree view with a depth dimension; the root folder is at the bottom edge of the screen, and the subfolders are above it. This gives the impression that the subfolders are behind the root folder. Results of the evaluation of the Space Manager showed that the spatial structure makes it easier to perceive its contents on a mobile phone (Hakala, et al., 2005). The results of previous research are contradicting, but it must be emphasized that these studies were focused on file and document management. Therefore, they used more files (up to 99 web page images) and more complex filing systems. This might have lead to the result that 3D interfaces were more cluttered. In our research, experiments will be done that are focused on icons in a menu interface of a mobile device, which means that the amount of icons will be restricted. An experimental psychology study indicated that in a visual search task depth can be processed very efficiently (Geib & Baumann, 1990) and it can also reduce the effects of visual crowding (Felisberti, Solomon, & Morgan, 2005). Hence, depth can give icons visually more space, since icons can be positioned on different depth layers instead of just on a row or in a grid.

A side effect of using stereopsis is that images can become blurred because of crosstalk. Crosstalk causes the viewer to see partially adjacent views. There are ways to use this blurriness to the advantage of the design of the graphical user interface. For example, it is known that when people look at a particular object, the object is sharp but the background is slightly blurred, which can be compared to the depth of focus effect in normal viewing (M. Lambooij, IJsselsteijn, Fortuin, & Heynderickx, 2009; M. T. M. Lambooij, IJsselsteijn, & Heynderickx, 2007). It is possible to simulate this effect by making the object of interest on the graphical user interface sharp and the background somewhat blurred by rendering the background behind the display panel and the object of interest on the display. To evaluate whether this design affects the users' preference, we here report on experiments done with four different depth modes for the location of backgrounds and objects of interest. The four depth modes are called:

- Sinking background, where the background is behind the display and the object of interest on the display,
- Floating icon, where the background is on the display and the object of interest (in this case an icon) is hanging in front of the display,
- Window effect, where the background is behind the display and the object of interest is also behind the display, but in front of the background,
- On display, here both the background and object of interest are on the display (this is actually a 2D interface).

In literature one experiment is reported that assessed the effect of sinking background and floating text with four different backgrounds. In this experiment subjects had to search for words. They found that the sinking background had a significantly better performance than the floating text (Mizobuchi, et al., 2008). This might be explained by the fact that in the sinking background mode the text is sharper than in the floating text mode. Like in the experiment conducted by Mizobuchi, et al., this experiment will also use different depth modes and backgrounds.

1.3 Evaluation

Before a new technology such as an auto-stereoscopic displays for a mobile device can be brought to the consumer market it has to be tested and evaluated. It is important to understand the way users use the product and what they like about it. Additionally, it allows designers to make necessary improvements before the product actually appears on the consumer market.

To evaluate a user interface for an auto-stereoscopic display on a mobile device, two types of tests can be done, namely perception tests and usability tests. Perception tests investigate the visibility of e.g. noise, blocking artifacts, and luminance, but also how much a stimulus has to be changed until it is visibly different from a reference or standard. Additionally, perception tests can be used to measure the preference of a user to certain visual information, i.e. to what degree the user prefers a stimulus. The latter type of perception tests is used in this research as a means to evaluate the image quality of a stimulus, and with that the added value of an auto-stereoscopic display for the user interface on a mobile device. From this test it can be found for example how much disparity is preferred by the subjects.

It is important to not only test what the users will see on the user interface, but also how they see it from the perspective of using it. This can be tested with usability tests. Most usability tests are done by giving test subjects tasks to do with the user interface (Kaikkonen, Kekäläinen, Cankar, Kallio, & Kankainen, 2005). With such tests the error rate and time to complete the tasks are measured amongst other measurements (Shneiderman & Plaisant, 2009). For the experiments done in our research, however, there was no working prototype available. Therefore, no tasks could be given to the subjects. As a consequence, participants were asked to judge the affordance and aesthetics of images of user interfaces shown to them. Affordance is the easiness of detection of an action behind an object or environment. For example, a handle affords to be pulled or to be lifted; a button affords to be pushed (Dix, Finlay, Abowd, & Beale, 2003). In the experiments described here, participants had to look at the affordance of the icon. The aesthetics of the user interface is the visual attractiveness and appearance.

Perception and usability tests can be combined in various ways. In addition to these combined tests, user surveys can be used. These are familiar and inexpensive in their use.

1.4 Problem definition

For a new technology like an auto-stereoscopic display in a mobile device, some questions have to be answered. One of them is how an auto-stereoscopic display can create added value in the design of a user interface for a mobile device. Additionally, it is interesting to see how the 3D effect induced by stereopsis can be enhanced by using existing visual depth cues, such as shadowing, luminance differences, and size differences. Hence, this research focuses on the following question:

“When using an auto-stereoscopic display on a mobile phone, which depth cues can be used on the user interface to increase its overall experience?”

To be able to answer this question, two sub-questions have to be addressed.

It is not clear yet what the important attributes will be to answer the research question. Do the users look at the colors, do they look at the artifacts or do they look more at the easiness of use. The most important attributes for users to assess the quality of a graphical user interface have to be found first. Therefore, the following sub-question has to be answered as well:

“Which attributes do users of mobile devices use to assess a graphical user interface of a mobile device?”

To answer this question three attributes are investigated in our research, namely *affordance*, *aesthetics*, and *preference*. These attributes will be more extensively discussed in Chapter 3.

Besides the question which attribute is important, there is also the question, whether monocular depth cues can help to give an optimal impression of depth organization, and with that a more appreciated overall impression. A selection of depth cues will be used based on their usefulness, which needs to be evaluated in a separate experiment. This separate experiment addresses the second sub-question, which is:

“Which depth cues can be used to enhance the depth perception?”

To answer all three questions three experiments will be done.

1.5 Outline of this thesis

This thesis is organized as follows. In chapters 2 and 3 literature on related topics is described and used to formulate the hypotheses, which are defined in chapter 4. In chapter 5 the pilot experiment, where the three attributes for assessing a graphical user interface are investigated, is described. The tuning experiment, which investigates the usefulness of the depth cues, is described in chapter 6 and the main experiment, where the research question will be answered, is described in chapter 7. Conclusions and future works are given in chapter 8.

2 The third dimension

Humans see depth in daily life by using visual and oculomotor cues. There also exist illusory depth effects; they create a feeling of depth, but it actually is only an optical illusion. When displaying images and videos on 2D displays, an impression of depth is perceived by using visual cues. But nowadays display technologies exist that render real depth, by using stereoscopy. The various techniques to create stereoscopy can be classified in (1) stereoscopic displays using glasses as a viewing aid like in the movie theaters, and (2) auto-stereoscopic displays, which do not need viewing aids.

This chapter first describes in section 2.1 the visual and oculomotor cues, and then in section 2.2 the illusory depth effects. Section 2.3 describes the various 3D display technologies. Watching 3D images and videos can cause discomfort, which is discussed in section 2.4. To conclude this chapter, formats for displaying stereoscopy are discussed in section 2.5.

2.1 Depth cues

Humans have two eyes that are located closely side-by-side at the front of their heads. Because of this position both eyes get a slightly different view of the same scene. The differences in the two views allow the brain to extract depth information. This phenomenon is called stereopsis. The differences in the two retinal images are lateral displacements, which are also called disparities. It is, however, not strictly necessary to have two eyes to get an impression of depth. Monocular cues are based on one eye only and can give some sense of depth as well. Actually, these are the cues to help us understand depth positioning in 2D images and video. The following monocular cues for static images exist (Ware, 2004):

- Occlusion: an object that partly hides another object is closer to the viewer.
- Linear perspective: parallel lines converge to each other at a large distance.
- Aerial perspective: distant objects are slightly blurred and bluer than objects near the viewer, because light scatters over distance and blue light scatters more than green or red light (Cockburn & McKenzie, 2004).
- Texture gradient: texture elements decrease in size with distance.
- Size gradient: an object with a known size gives a scale to the whole picture.
- Shadows: give information about the shape of the object and help to locate the object with respect to a surface in the image.
- Depth of focus: objects nearby and farther away from our point of focus become blurred.

Monocular depth cues can be powerful to give a sense of depth (Heynderickx, 2007). Occlusion is reported as the most dominant monocular cue (Cutting & Vishton, 1995). Blur variation is also reported to render an apparent depth ordering (Mather & Smith, 2002). But monocular cues are not always very accurate in giving a sense of depth. In a research it was shown that illumination inconsistencies do not 'pop-out' (Ostrovsky, Cavanagh, & Sinha, 2005). This means that various objects in a scene can have different shadow directions without it becoming very obvious.

A powerful dynamic cue is motion parallax. It refers to the phenomenon that objects close by move faster than objects farther away (Heynderickx, 2007). Sometimes, also the term structure-from-motion is used to describe this phenomenon (Ware, 2004). This is an overall term including both motion parallax and the kinetic depth effect. The kinetic depth effect is also a dynamic cue, describing the phenomenon that a non-moving object appears to be two dimensional, but its three dimensional shape becomes immediately apparent when it starts to move.

To really accurately see depth layering, stereopsis is needed, and so two eyes are needed. Therefore, stereopsis is referred to as a binocular cue. Stereopsis is actually the only binocular cue.

Next to the monocular and binocular visual cues there are also oculomotor cues. Oculomotor cues provide a sense of depth via the muscular activity in the eyes. There are two oculomotor cues (Heynderickx, 2007):

- Accommodation: the changing of the shape of the lens to keep an object of interest in focus at the retina.
- Convergence: the muscular strain, resulting from the eyes rotating inwards for objects close by and outwards for objects farther away, to keep the object projected at the centre of the retina.

2.2 Illusory depth effects

Besides real cues there are also effects that give an illusory depth effect. One of these effects is chromostereopsis. It refers to the illusory depth effect where for example red letters appear to stand out of a black background while blue letters have a reversed effect for most people. Chromostereopsis is caused by chromatic aberration (Ware, 2004).

Another effect is the "pop-out" effect caused by differences in brightness (achromatic pop-out) (Theeuwes & Lucassen, 1993). It has been shown that a brighter object embedded in an array of identical other objects on a dark background "pops-out". However, it is also suggested that "pop-out" is not a characteristic property of luminance itself, but it is dependent on the set-size of the objects (Baldassi & Burr, 2004). The effect of difference in luminance is also sometimes called a "theatrical" effect (Cutting & Vishton, 1995), as it is assumed that in a theater the foreground is better lit than the background.

2.3 3D display technology

To display stereoscopic depth, specially designed displays are needed. Normal displays can only give an impression of depth by means of monocular cues. There are several kinds of displays that can render stereoscopic depth. These displays generate stereopsis, which implies that the left and right eye receive a slightly different view of the same scene. In order to do so, disparities are included in the content in such a way that they yield the appropriate retinal disparities. The oldest way to generate stereopsis is by using viewing aids like glasses; these methods are discussed in section 2.3.1. Nowadays, technology has advanced so far that there are displays that are capable of generating stereopsis without viewing aids. These displays are called auto-stereoscopic displays. There are two kinds of auto-stereoscopic displays, i.e. flat displays and volumetric displays. For the flat displays there are two options: they can be either two-view or multi-view. The two-view displays render stereoscopic depth only at one specific viewing position; unless head-tracking is applied they provide no look-around effect. With multi-view displays viewers can walk up and down in front of the display and still see stereoscopic depth. Hence, these displays can provide the look-around effect. Auto-stereoscopic displays are discussed in section 2.3.2.

2.3.1 Stereoscopic displays

There are several display techniques for stereoscopic displays. The following techniques require the viewer to wear glasses (Dodgson, 2005):

- Colored glasses, anaglyph
- Polarized glasses
    - Two standard displays, made coplanar by a half-silvered mirror

    - Two projectors, projecting onto a polarity-preserving screen
- Shuttered glasses, combined with a display with a double-frame rate

In the next sections these techniques are described.

2.3.1.1 Anaglyphs

The anaglyph method is an inexpensive way of providing stereo to large groups of people. In this method, the difference in left and right eye can be realized in different ways; the simplest way is by using a contrasting color filter for each eye. This has to be combined with a projector displaying the image for each eye in the same contrasting colors. When looking through the color filters 3D can be seen. This anaglyph method is relatively easy to make, and crosstalk can be largely avoided if the color filters are properly designed. The disadvantage of this method is its poor color rendering of 3D images. This is caused by the coupling of depth information to color information.

2.3.1.2 Polarized glasses

The principle of creating stereopsis with polarized glasses is almost the same as for the anaglyphs. The color filters are here replaced by two polarizers. The polarization direction is mutually perpendicular, for example the polarizer for the left eye is vertically oriented, while the polarizer for the right eye is horizontally oriented. As a consequence, the projector showing the left image needs a vertically oriented polarizer, while the projector showing the right image needs a horizontally oriented polarizer. Compared to the anaglyph method, the method using polarized glasses has a much better color rendering. Also crosstalk can be kept small when good polarizers are used. The disadvantage of using polarizers is that they block half of the light intensity, and therefore, result in a low brightness.

2.3.1.3 Shutter glasses

Besides using color filtered and polarized glasses it is also possible to use shutter glasses. The shutter glasses alternately close for the left and right eye, while keeping the other eye's view open. The display shows time-sequentially the left and right eye's view of the scene in perfect synchronization with the shutter glasses. This technique can only work when the synchronization is very accurate between the glasses and the display for a frequency that is sufficiently high to prevent perceived flicker.

2.3.2 Auto-stereoscopic displays

As mentioned above stereoscopic displays make use of visual aids. For a mobile device, however, it would not be very practical for a user to carry around an extra pair of glasses to use the stereoscopic mode of the mobile device. Auto-stereoscopic displays do not require visual aids, and therefore, these kinds of displays are more practical for mobile devices. There are various kinds of auto-stereoscopic displays, which are described in the following sections.

2.3.2.1 Two-view displays

A two-view display produces two images, i.e. one for the left eye and one for the right eye. Because there are only two views, the user has to sit in a fixed position in front of the display, so that his left eye catches the image for the left eye and the right eye catches the image for the right eye. These systems are relatively simple and inexpensive, and because the light is usually well separated for both eyes, there is limited to no crosstalk. The loss of spatial resolution is only a factor of 2, as there are two pixels of the original display resolution needed per 3D voxel to create the two views. The depth range of these displays is limited and they suffer from accommodation-convergence rivalry (this will be explained in more detail in section 2.4). With a two-view display it is not possible to have a look-around effect; there is only one perspective of the 3D scene displayed.

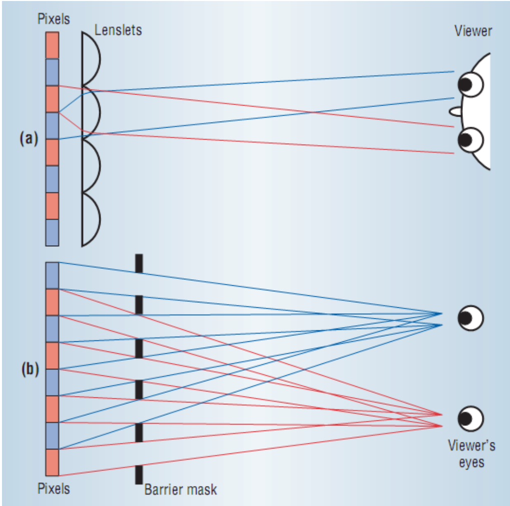


Image copied from (Dodgson, 2005)

Figure 2.1 The two ways to create a two-view display using in (a) a lenticular and (b) a parallax barrier

A parallax barrier is most often used to separate the two views for a two-view auto-stereoscopic display (see Figure 2.1b). A parallax barrier consists of a series of slots that are accurately aligned with the underlying pixel structure of an LCD. The light that comes from the pixels containing information for the right eye is blocked for the left eye by the barrier, and the light with information for the left eye is blocked for the right eye. Another way of creating a two-view display is using a lenticular sheet (see Figure 2.1a); this consists of an array of cylindrical lenses which is placed in front of a LCD display. The lenses direct the light coming from the pixels to different directions creating different views, this way each eye receives light from only every second pixel column. When the user has his eyes in the dedicated viewing zones a nice 3D image can be seen, but when one eye exceeds the viewing zone, crosstalk will be perceived or image information will be missing.

2.3.2.2 Head-tracked display

To have a look-around effect with a two-view auto-stereoscopic display head-tracking is necessary. There are two ways of adapting the 3D display to head-tracked information. The first way is applied in a 3D display with multiple cones of two views, as can be seen in Figure 2.2a. The viewing zones are swapped at the display when the head moves such that the right eye is in the left-eye's viewing zone and the left eye in the right-eye's viewing zone. The swapping ensures that each eye is in its appropriate viewing zone. This swapping approach can only be applied for one viewer at a time. Another way of adapting the 3D display to head-tracked information is used for 3D displays that render two views in only one specific direction, as can be seen in Figure 2.2b. For these displays, the viewing direction is adapted to the position of the head based on the head-tracked information. This method can be used for more than one viewer, but then the display must be capable of displaying multiple sets of two views.

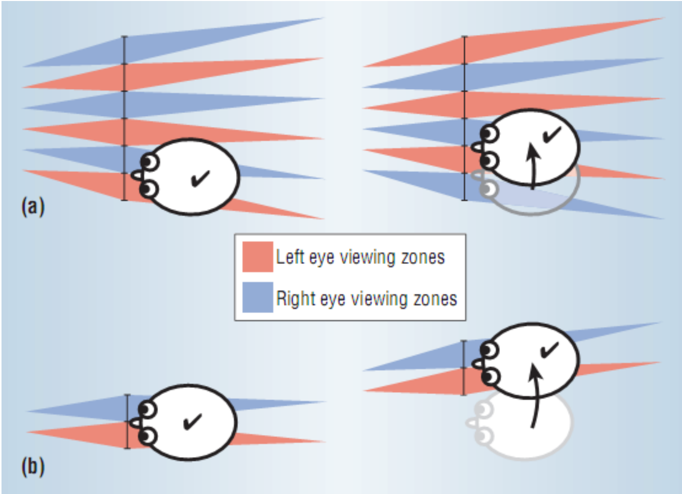


Image copied from (Dodgson, 2005)

Figure 2.2 Two ways of two-view head-tracked displays

2.3.2.3 Multi-view display

With a multi-view display the viewer can move in front of the display while still perceiving 3D. The only condition is that the eyes have to capture two different views. A multi-view display shows multiple cones with three or more views per cone. These displays are relatively easy to make and can accommodate more than one viewer. But, as for most stereoscopic displays, these displays suffer from accommodation-convergence rivalry (see for a more detailed explanation in section 2.4), and they can only display a limited field of depth. Additionally, multi-view displays suffer from a large loss in spatial resolution as the number of sub-pixels has to be divided by the number of views.

Multi-view displays can be made using a parallax barrier, but this is not the best approach as the luminance decreases with the number of views. A better alternative is using a lenticular sheet generating multiple views.

The effect of the number of views used on the viewing experience with auto-stereoscopic displays was studied (Salmimaa, Häkkinen, Liinasuo, & Järvenpää, 2009). Here two-view, five-view and fourteen-view 3D displays were mutually compared. The results indicated that the two-view was significantly better than the five-view or fourteen-view regarding readability and image sharpness.

2.3.2.4 Volumetric displays

In volumetric displays there is no accommodation-convergence rivalry (see for a more detailed explanation in section 2.4), as the image is produced within a volume of space, where the space is either real or virtual. Light is generated directly at the exact depth where the image information is situated. Because light that is generated at a deeper layer cannot be blocked by other layers closer to the viewer, images of a volumetric display are always translucent.

2.3.2.5 Holography

A hologram can be considered as a display with infinite number of views, where the viewer receives a separate image for both eyes without any crosstalk and with full motion parallax. An image from a holographic display is reproduced by wavefront reconstruction. The exact waveform as in a real scene is reproduced at each voxel. This technique does not suffer from accommodation-convergence rivalry. However, this is a complex system and not feasible for video as exact wavefront production is not possible with natural lighting, but only with monochromatic light.

2.4 Visual discomfort and visual fatigue

There are numerous ways to enhance the user experience with a stereoscopic display: the 3D display can enhance the entertainment experience, but can also be used to improve the user interface or visualization of data (Kooi & Toet, 2004). However, using stereoscopic displays can also create problems, resulting in visual fatigue and visual discomfort. Visual fatigue refers to a decrease in performance of the human vision system, which can be objectively measured, whereas visual discomfort is its subjective counterpart (M. Lambooij, et al., 2009). There are four groups wherein visual fatigue can be divided in (M. T. M. Lambooij, et al., 2007):

1. Asthenopic – eyestrain, tired and sore eyes, feeling of pressure in the eye and chemical changes in intracorporeal substances,
2. Ocular surface-related – dried mucus, painful irritation, tearing, reddening of the eyes and conjunctivas,
3. Visual – double vision, blurred vision, slowness of focus change, reduced sensitivity to spatial contrast, visual acuity and speed of perception, reduced power of accommodation and convergence and presbyopia,
4. Extra ocular – headaches, ache around the eyes, neck pain, back pain, and shoulder pain, distortions of psychological activities in humans and subjective symptoms such as a decline in work efficiency and loss of concentration.

One of the possible causes of visual fatigue is the amount of depth. Even though it might be tempting to use large depth values to maximize the viewing experience, it is not advisable, since too much depth can exceed the human ability of fusing images. As a consequence, diplopia can occur, which refers to seeing double images instead of one image. Eye movements are known to decrease diplopia, which thus leads to larger tolerances for depth ranges. This means that when images are shown for a longer period in time, they can have more depth, but images that disappear quickly should use smaller depth ranges (Häkkinen, Takatalo, Kilpeläinen, Salmimaa, & Nyman, 2009). By increasing the depth gradually, it is also possible to use depth ranges for a short period of time. When it is used for a longer period, eye strain still becomes a problem (Häkkinen, et al., 2009).

Apart from the depth range, also the accommodation-convergence rivalry is mentioned as a possible cause for visual discomfort (M. Lambooij, et al., 2009). As explained before, accommodation refers to the ability of the human eye to adapt the lens thickness to keep the object of interest in focus at the retina. And convergence refers to the rotational eye movement to keep the object projected at the centre of the retina. In natural viewing conditions a change in accommodation activates a change in convergence, and any change in convergence activates a change in accommodation. However, in most stereoscopic displays this is not the case. Since the image is displayed sharpest at the distance of the display the eyes tend to accommodate there, while the eyes tend to converge at the distance of the object hanging in front or behind the screen. This is known as the accommodation-convergence rivalry. Accommodation-convergence rivalry can cause blurred images and double vision (Hoffman, Girshick, Akeley, & Banks, 2008; M. Lambooij, et al., 2009; M. T. M. Lambooij, et al., 2007). The accommodation-convergence rivalry is often said to be a major cause for visual discomfort when watching a 3D display. But the depth of focus for 3D displays that only show reasonable disparity values should be enough to avoid accommodation-convergence rivalry (M. T. M. Lambooij, et al., 2007).

A third possible cause of visual fatigue is crosstalk, or more generally unnatural blur. When information for one eye leaks into the other eye (interocular crosstalk) depth perception can be degraded (Yeh & Silverstein, 1990). However, crosstalk may also have a beneficial effect on image quality and visual comfort. Some auto-stereoscopic multi-view displays induce crosstalk to avoid a picket-fence effect (banding) and to minimize image flipping (M. Lambooij, et al., 2009).

High-level cue conflicts are not discussed yet too much as possible cause for visual fatigue. These conflicts occur when immersive displays attempt to re-create real-world scenes with various depth cues which convey different magnitudes of depth. High-level cue conflicts create a form of mental strain, because the brain tries to make sense of the incoming conflicting perceptual information (Patterson & Silzars, 2009).

Interocular luminance differences can be regarded as a particular case of a high-level cue conflict. Amounts greater than 60% of difference in luminance affect the depth perception in stereoscopic displays (Boydstun, Rogers, Tripp, & Patterson, 2009). The presence of unwanted vertical disparity is another cause of a high-level cue conflict. Because of the horizontal position of our eyes, our brain is not accustomed to vertical disparity (Kooi & Toet, 2004).

In summary, the most likely causes of visual discomfort are one or more of the following factors (M. T. M. Lambooij, et al., 2007):

1. Excessive demand of accommodation-convergence linkage, e.g. by fast motion in depth viewed at short distances,
2. 3D artefacts coming from insufficient depth information in the incoming data signal producing spatial and temporal inconsistencies, and possibly generating high-level cue conflicts,
3. Unnatural amounts of blur, e.g. caused by crosstalk.

To adequately understand visual discomfort and visual fatigue more research is still needed.

2.5 Formats

There are two different video formats in which stereoscopic content can be provided to a display. The first format consists of separate images for both eyes. This format represents the normal way humans perceive depth. This format is most useful for two-view (auto-)stereoscopic displays, since the incoming signal contains all data needed. But, the format can also be used in a multi-view auto-stereoscopic display. If such a display has nine views, nine different images have to be generated out of the existing two images. These nine images are then usually sent to the display as interleaved into one image.

The second format is called the 2D + depth format. This format consists of a 2D image with a corresponding depth map. The depth map is made up of grey levels, which represent the disparity. When an object in the 2D image has a corresponding grey level value of 0 in the depth map, then this object is displayed at the maximum distance behind the display panel. An object with a corresponding grey level of 255 in the depth map is displayed at the closest distance near the viewer. For this format all 3D information is given in the depth map, being just a grey scale image. As a consequence, broadcasting or storing the 3D data in the 2D + depth format requires less bandwidth than in the stereoscopic format discussed above, in which two full-sized images have to be broadcasted or stored. Besides this advantage, the 2D + depth format has two other important advantages. The first one is that all the views (depending on the display) are easily generated from the depth map. The second advantage is that the depth range is easily adaptable by adjusting the gain and offset of a display. The gain determines the maximal amount of depth that can be displayed in the total depth range before and behind the display screen. The offset determines how this depth range is positioned with respect to the display screen, i.e. more of the depth range in front or bind the screen.

In this research two displays are used. The format that is used for the display of the pilot experiment is the 2D + depth format. The display used in the tuning experiment and main experiment require an input image that consists of four quadrants. The top two quadrants are a 2D image and the corresponding depth map. The bottom two quadrants are used for occlusion information, i.e. the content that is occluded in 2D, but needed for a look-around effect in 3D. Also for these two quadrants a 2D image and a corresponding depth map is used. Only now the information of the occluded objects are given in the 2D image and depth map. When occlusion information is not used, the bottom two quadrants can be left bank.

3 The user interface

User interfaces are used on almost all electronic devices, such as DVD-players, televisions, computers, mp3 players, mobile phones…etc. The user interface is the means for humans to interact with the electronic equipment. These interactions can be as simple as adjusting the volume or as complicated as programming a series of actions on a computer. Obviously, the diversity in interaction complexity has led to a diversity in interaction styles, which are discussed in section 3.1.

It is essential that the user interface is easy to use. Therefore, it has to comply with some golden design rules, which are summarized in section 3.2. This section also describes a design process that aims to optimize the chance of success of a user interface in terms of usability.

Most of the design rules and processes as reported in literature for a user interface have a computer monitor application in mind. Because of obvious differences, the user interface design of a desktop computer cannot directly be applied to a mobile device. The functionality on a mobile device is more limited than on a desktop computer, a mobile device uses different input means than a desktop computer (e.g. a touch screen rather than a mouse), the display of the mobile device is smaller and the power of its CPU and memory is more limited. Section 3.3 discusses these differences and their consequences for the design of the user interface in more detail.

Section 3.4 then describes the addition of 3D in the user interface of a mobile device.

After being designed, the user interface needs to be evaluated on its functionality and aesthetics. Methods that can be found for these evaluations are discussed in section 3.5.

3.1 Interaction styles

A user interface can make it easier for a user to interact with a system on conditions that the interaction style is appropriate for the required functionality. As mentioned above, there is a wide range of interactions that users may want to have with a system. As a consequence, there are various interaction styles (Dix, et al., 2003). The most common styles are:

- Command line interface – this interaction style requires users to type in specific commands. Therefore, it is more used by expert users, as it is necessary to remember the commands. Nowadays it is mostly used in addition to menu-based interfaces.
- Menus – in this interaction style the available options are displayed on the screen, and can be selected by using the mouse, or numeric or alphabetic keys. This interaction style is less demanding for the user, since the options are visible to the user; the user only has to recognize the option and not recall it as with the command line interface.
- Natural language – in this interaction style users give commands to the interface in spoken language. Natural language, however, is difficult to understand for a machine because of its ambiguity. Where humans rely on context and general knowledge when interpreting spoken language, machines are not yet able to do so sufficiently accurately. Therefore, it is unlikely that a system will have an interface that uses general natural language, but a system that understands a restricted subset of a language could be build.
- Question/answer and query dialog – in this interaction style the user is led through the interaction by questions, which require mainly yes/no, multiple choice and code answers. This interaction style is easy to learn, although effective use of query language requires some experience.

- Form-fills and spreadsheets – this interaction style is primarily used for data entry, but can also be used for data retrieval applications. The user has to fill in a form, which resembles a paper form, but is presented on the display. In a spreadsheet users are free to manipulate values. This makes the distinction between input and output less visible and the spreadsheet interface more flexible and natural.
- WIMP – this interface style is currently used by the majority of computer systems. WIMP is an acronym for windows, icons, menus and pointers. This style is often called a graphical user interface (GUI).
- Point and click – this interface style refers to every system that uses a mouse interface or touch screen interface, with which users can point to a given action and confirm the action by clicking on it.
- Three-dimensional interfaces – three-dimensional effects in user interfaces are for example used in virtual reality, but simpler versions are also used in WIMP elements. The elements are given a three-dimensional appearance by using shading. Interfaces with 3D workspaces are in general more complex; the objects displayed are usually flat, but the objects are displayed in perspective when the user is looking at it from an angle and smaller when they are further away from the user. Here size, light, and occlusion give a sense of distance.

The decision on which interaction style to used does not only depend on the required functionality, but also on whether the target group consists of expert or novice users. It is needless to say that an expert will not be pleased with a user interface built for a novice user. The expert prefers to use shortcuts to work more efficiently or to use the command line interface to use the full flexibility of the electronic equipment. A novice user, on the other hand, needs a simpler interaction style with more background information. It is a real challenge to design a user interface that is easy and pleasant to use for all types of users.

3.2 Guidelines for user interface design

According to (Redmond-Pyle & Moore, 1995) a user interface consists of:

- The computer hardware which allows the user to interact with the system; this includes, for example, the keyboard, mouse, buttons and screen,
- The image material on the screen, which includes, for example, the windows and menus,
- The user documentation, such as manuals.

For the development of a user interface there are four steps defined that increase the chance of developing a successful user interface, called the four pillars (Shneiderman & Plaisant, 2009). These four steps consist of:

1. User interface requirements – the requirements of the user interface have to be gathered. This should be done as precise and complete as possible. One way to determine the user interface requirements is to use ethnographic observations. With an ethnographic observation the users are observed in their own environment.
2. Guideline document & process – a document should be made that contains a set of working guidelines. This way all the developers use the same design principles. When such a document exists the implementation proceeds more quickly.
3. User interface software tools – various software tools are available to implement a user interface in. It is important to use the tools that are suitable for the specific user interface that needs to be designed. Also it is important to make a prototype before actually implementing the user interface. For the prototype it is preferred to have some kind of interaction with the keyboard

and mouse, since that makes the user interface more realistic and gives the user an impression of the final system.

4. Expert reviews & usability testing – before releasing the final product it is important to do various tests. Conducting these tests allows the designers to find the errors and correct them before release. There are various ways of testing; in the next section some of these methods are discussed in more detail.

The four steps do not guarantee a flawless working system, but it does support the developer on where to start when designing a user interface for a given system. A golden rule for designing good user interfaces is to "understand your materials" (Dix, et al., 2003). For human-computer interaction the materials are:

- The human – understand the human's psychological, and social aspects, and how humans make errors,
- The computer – understand the limitations, capacities, tools, and platforms of the system.

Apart from this general design rule, various golden rules for user interface design have been reported in literature. For the user interface of a desktop computer the "Golden Rules of Interface Design" of Shneiderman are often referred to. They define eight design principles, which are (Shneiderman & Plaisant, 2009):

1. Strive for consistency – in similar situations consistent actions should be required; identical terminology should be used; consistent color, lay-out, fonts, and so on should be used throughout the user interface.
2. Cater to universal usability – various users should be able to use it, whether they are novice or expert, young or old, have disabilities or not. The content should be able to be transformed to the need of the user.
3. Offer informative feedback – the system should provide feedback for every action that has been activated by the user.
4. Design dialogs to yield closure – action sequences should be organized into groups with a beginning, middle and end.
5. Prevent errors – the system should be designed in a way that it is almost impossible for the user to make a serious error. Good examples are disabling buttons that cannot be used at that moment in time or deactivating alphabetical characters in numerical entry fields. When an error is detected, the interface should provide simple, constructive, and specific instructions for recovery.
6. Permit easy reversal of actions – actions should be reversible as much as possible.
7. Support internal locus of control – experienced users desire a sense of control of the interface. They do not want to be surprised by the interface.
8. Reduce short-term memory load – interfaces that require users to remember information from one screen and use that information on another screen should be avoided.

3.3 User interface for mobile devices

There are essential differences between a desktop computer and a mobile device that will affect the user interface as well. The functions on a mobile device are mainly limited to calling, simple text messaging, and editing an address book. The more sophisticated mobile devices also allow the user to contact to the internet and read e-mail, but, even this functionality is more restricted than on a desktop computer. Additionally, the memory capacity and processing power is smaller on a mobile device than on a desktop computer. These capacities are continuously growing, but nonetheless, remain considerably smaller for the mobile than for the desktop device. Also the display size of a mobile device is much smaller than that of a desktop computer. Even though the display of a mobile device has grown considerably in size during the last years, this will not continue much farther, since there is a natural limit to the display size of a mobile device. Finally, the input tools are different for a mobile device than for a desktop computer. Where the desktop computer typically uses a keyboard and a mouse, the mobile device usually uses only a limited keyboard design. Again, the more sophisticated mobile devices have higher level means of input, being a full keyboard and a touch screen, but because of the size of the keys on the keyboard, typing in commands, for example, is more inconvenient on a mobile device than on a desktop computer.

These differences imply that the interface of a desktop computer cannot be simply copied to a mobile device. As an example, this is clearly illustrated by the difference in display size. The larger display on a desktop computer can show a lot of information while staying uncluttered, but the small display on a mobile device with the same amount of information will look cluttered and will not at all be user friendly. Apart from the actual interface, also the interaction styles and guidelines for the user interface are different for a mobile device and desktop computer. Contemporary mobile devices mostly have an interaction style based on a menu user interface paradigm. To avoid derogating the usability in mobile phones, ambiguous naming and allocation of functions in the menu should be avoided (Bay & Ziefle, 2005). The typical interaction styles for mobile phones remain reasonably stable; users who are transferring from their old phone model to a new phone model do not face significant difficulties (Kiljander, 2004).

The guidelines given for a user interface in section 3.2 were more applicable to personal computers than to mobile devices. Due to the limitations of a mobile device as discussed above, some of these guidelines have to be adapted (Gong & Tarasewich, 2004). Users might need to switch between their desktop computer and a mobile device, for example to transfer documents to their PDA. This adds another dimension to the basic rule of consistency. Because of the lack of available resources, the reversal of actions can be more difficult for mobile devices than for desktop computers. Errors should be prevented on a mobile device as on a desktop computer, but due to the physical design of the mobile device the errors can be of a different nature; the small buttons of the device can cause the user to press the wrong buttons. Users of mobile devices have likely more distractions, and therefore, the interface should be designed in a way that very little memorization is necessary for performing tasks. Besides these adaptations the following challenges for the user interface of a mobile device exist (Dunlop & Brewster, 2002; Gong & Tarasewich, 2004):

- Designing for mobility – the working environment of the users changes continuously as users move; the diverse environmental conditions, such as sunlight, noise, and weather, have to be taken into consideration,
- Designing for a widespread population – since the user interface should be understandable and usable for every user, personalization of the device is important. The user interface should also be designed to offer enjoyment,
- Designing for limited input/output facilities – because of the factor of portability, screen sizes will not exceed a specific size. Keyboards are very small and often limited in the number of keys. Due to these physical limitations, new interaction techniques may be necessary,
- Designing for multitasking – users of mobile devices are often focusing on more than one task, for example driving a car while the mobile phone rings. It is important that a mobile device is designed in a way that the interface needs as little attention as possible.

3.4 3D user interface for a mobile device

As mentioned earlier in the report, 3D displays, rendering content stereoscopically, have been introduced to the market. Since also for mobile devices the display shows an improvement in resolution, color gamut and computational power, the introduction of some kind of 3D technology becomes feasible on these devices as well. With a 3D display the mobile device can give added value to the gaming experience (Rajae-Joordens, 2008) or can de-clutter the menu structure by using various depth planes.

The Sharp mova SH505i has been used in a study to investigate the benefits of a stereoscopic display. In this study participants had to look for a particular word. The words were on top of the background (Mizobuchi, et al., 2008). This study consisted of two experiments. In the first experiment participants had to look for text on (1) a normal 2D display, (2) a 3D display based on a parallax barrier, and (3) a 2D display of equivalent brightness to the 3D display. Participants were significantly faster in finding the word on the normal 2D display than on the 3D display, but this advantage in speed disappeared when the 2D display had the same brightness as the 3D display. In the second experiment, they compared four background conditions varying in where the text and background were positioned in depth (e.g. floating text against sinking background). They found that for a parallax barrier 3D display the layer of primary interest should be displayed with zero disparity, while the secondary layer should appear sunken behind the display screen.

This layering can also be used to de-clutter the icons. Icons of interest can be put on the top layer. The rest of the icons can be put on the background, which can be rendered behind the display screen. Whether this layering improves the user interface is further investigated in this report.

3.5 Evaluation

When 3D is added to the user interface of a mobile device it adds a dimension to the evaluation. It is not only important that its usability is good, but also its viewing experience is important. Therefore, two kinds of tests have to be done; a perception test and a usability test.

3.5.1 Perception testing

Visual perception testing can in general be divided into two categories: visibility and preference testing. Visibility testing measures the threshold at which a given aspect is visible. Two types of thresholds can be distinguished: the absolute threshold and the just-noticeable difference (JND). The absolute threshold indicates the minimal value of a physical stimulus that is just visible or detectable by an observer. Examples can be the visibility of noise or blocking artifacts. Related to 3D displays, relevant absolute thresholds are the amount of disparity needed to have stereoscopic vision, or how much crosstalk can be introduced in a 3D display before it is visible. A JND indicates the minimum extent of change in a stimulus needed to perceive the stimulus as different from a reference. An example getting lots of attention in the display world is the visibility of color differences. To find an absolute threshold or JND, the staircase method or tuning method are commonly used.

Preference testing measures the degree to which a stimulus is preferred. Preference can be tested more specifically with the attribute image quality, or with the attributes viewing experience or naturalness when evaluating 3D displays (Seuntiëns, et al., 2005). The direct scaling method is most often used to assess preference. In that case, observers are asked to give a score referring to the attribute under test to each stimulus.

3.5.2 Usability testing

A user interface makes a system usable to the user, which makes the usability of a user interface the most critical aspect (Redmond-Pyle & Moore, 1995). Therefore, evaluating a system on its usability is important. According to ISO 9241 usability is:

"The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use."

In this definition effectiveness, efficiency and satisfaction are defined as:

- Effectiveness – "Accuracy and completeness with which users achieve specified goals"
- Efficiency – "Resources expended in relation to the accuracy and completeness with which users achieve goals"
- Satisfaction – "Freedom from discomfort, and positive attitudes towards the use of the product."

The principles of usability can be divided into three categories (Dix, et al., 2003):

- Learnability – the ease of understanding for novice users how to use the system and how to achieve a maximal level of performance.
- Flexibility – the variety of ways used by end-user and system to exchange information
- Robustness – features that supports the user to determine successful achievement and assessment of the goals.

A number of methods for usability evaluation exist. They can be divided into expert reviews and usability testing. According to (Dix, et al., 2003) the following techniques can be used for an expert review:

- Cognitive walkthrough – the user interface is evaluated by going through some common tasks that the user would perform and by evaluating the support the interface provides,
- Heuristic evaluation – several usability experts evaluate the user interface design by applying a set of heuristics or design guidelines,
- Model-based evaluation – evaluation of the user interface using a model, e.g. dialog models can be used to evaluate dialog sequences for problems.

Several sets of usability heuristics were compared to existing usability problems in order to determine the heuristics that explain usability problems best (Nielsen, 1994). Some of the heuristics that followed from the comparison were aesthetics and minimalistic design. Aesthetics is part of designing an overall enjoyable user experience with the user interface. It is also defined as a dynamic interaction that invokes a positive response from the user (Karlsson & Djabri, 2001).

The previous list of expert review methods can be extended by the following (Shneiderman & Plaisant, 2009):

- Guidelines review – it is checked whether the user interface is conform to the guidelines document,
- Consistency inspection – the user interface is verified on its consistency in fonts, terminology, color schemes, and layout. This check also includes the documentation and online help,
- Metaphors of human thinking – the experts focus on how users think when interacting with the user interface,

- Formal usability inspection – a meeting is held with a moderator. The merits and weaknesses of the interface are discussed in this meeting. Design-team members are allowed to be there to defend their design choices.

A small study on the usability of mobile phones was done with nine subjects using their own mobile phone. The subjects were asked to perform 26 tasks. In this study it was found that the subjects had no problem with doing the most common tasks like calling. Finding functions that were not frequently used, however, was problematic (Klockar, Carr, Hedman, Johansson, & Bengtsson, 2003). These difficulties could have been prevented with a clear menu design. Also navigation on most mobile phones is an issue. Whether 3D technology can help to circumvent these issues will be further evaluated in this research.

4 Hypotheses

As stated previously the main research question is:

"When using an auto-stereoscopic display on a mobile phone, which depth cues can be used in the user interface to increase its overall experience?"

To answer this question, the main research question has been divided into two sub questions:

- *"Which attributes do users of mobile devices use to assess a graphical user interface of a mobile device?"*
- "*Which depth cues can be used to enhance the depth perception?"*

Related to these sub-questions we have several hypotheses that have to be evaluated in this research. These hypotheses are:

Hypothesis 1. *Users have an increased sense of depth when shadows are combined with stereopsis.*
Shadows give information about an object's shape, rather than on its position in depth (Cutting & Vishton, 1995; Heynderickx, 2007; Ware, 2004). Nonetheless, shadows can also help to locate an object with respect to a surface in an image. This means that when a selected icon in the graphical user interface of a mobile device has some shadow, it may be easier to position this icon with respect to the other icons in the graphical user interface. This effect likely enhances the sense of depth.

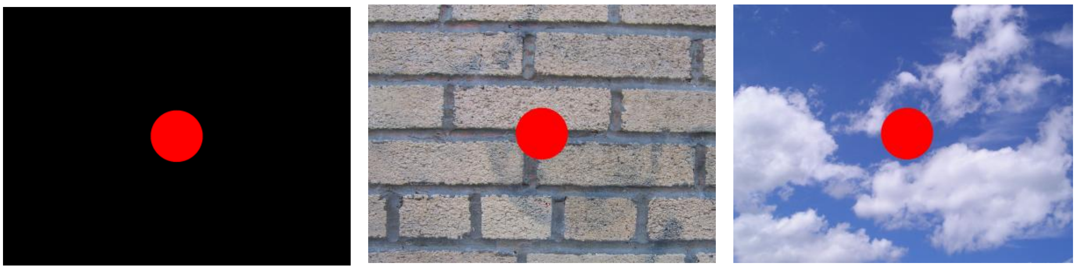
Hypothesis 2. *Users have an increased sense of depth when luminance differences are combined with stereopsis.*
Differences in brightness are known to create a kind of 'pop-out' effect (achromatic pop-out) (Theeuwes & Lucassen, 1993). Applying this 'pop-out' effect to a selected icon of a graphical user interface of a mobile device may give the user the feeling that the graphical user interface has more depth.

Hypothesis 3. *Users have an increased sense of depth when changes in size are used combined with stereopsis.*
It is known form monocular depth vision that an object with a known size gives a scale to the whole image. Here this phenomenon is used in the graphical user interface of a mobile device by changing the size of the selected icon. This is expected to create the impression that the selected icon is more to the front or coming out of the display.

Hypothesis 4. *Users have an even larger sense of depth when shadows, luminance differences, and changes in size are combined all together with stereopsis.*
If each monocular depth cue separately increases the sense of depth when combined with stereopsis, one would expect that the combination of all depth cues and stereopsis gives the optimal depth perception. The main question, however, is how these depth cues reinforce each other.

Hypothesis 5. *Users have a stronger sense of depth when using a natural instead of a uniform background.*

This hypothesis is based on the results of an experiment, in which a circle was displayed with different disparities on three different backgrounds: uniform black, a brick wall, and clouds (Arikan, Bos, & Wu, 2008), see Figure 4.1.



a) *b)* *c)*

Figure 4.1 The red circle on the three different backgrounds: a) the uniform black background,b) the wall, and c) the clouds

The disparities were chosen such that the circle came out of the screen, and the subjects were requested to indicate the location of the circle in depth with a cursor. According to the results the different backgrounds did not affect the perception of depth. However, participants indicated that it was harder to determine the location of the circle on the uniform black background than on the textured backgrounds. It could be that with the textured background the subjects could see that the circle was on top of a background, which means that the monocular depth cue occlusion was used to identify the location of the circle.

Hypothesis 6. *Users report a preferred experience when the selected icon of the graphical user interface is 'on the screen', while the other icons are placed 'behind the screen'.*

When the selected icon of a graphical user interface is at the screen and the remaining icons and background behind the screen, the icons at the background will be blurred to an extend depending on their disparity. This effect gives a similar impression as the depth of focus cue (which was explained in section 2.1). Note that a realistic depth of focus cue is difficult to be implemented, as it is not known where at the screen the user looks.

In (Mizobuchi et al., 2008) the reading performance of a situation where the background remains at the screen and the text comes out of the screen is compared to a situation where the text remains at the screen and the background is displayed behind the screen. This study shows that the reading performance is better when the text remains at the screen. Based on this result the authors propose that the layer of primary interest should be displayed with zero disparity, i.e. staying at the screen, while the secondary layer should be displayed with negative disparity, i.e. behind the screen.

Based on the known effect of the depth of focus cue and on the result of (Mizobuchi, et al., 2008), we expect a graphical user interface with the selected icon "on the screen to be preferred over other representations.

Hypothesis 7. *Users better appreciate the overall experience of a graphical user interface when they experience an increased sense of depth.*

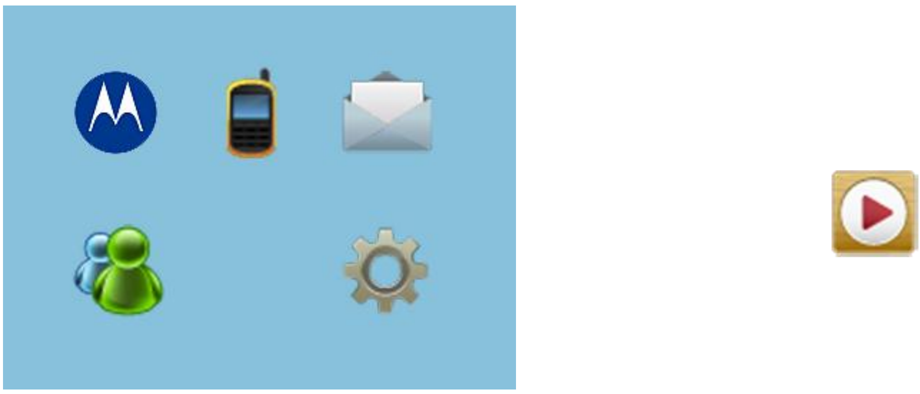
Literature shows that perceived depth can contribute to an enhanced feeling of presence and involvement (IJsselsteijn, de Ridder, Hamberg, Bouwhuis, & Freeman, 1998), and because of that, a stereoscopic display is more preferred than a 2D display. Depth can also help to navigate in a user interface (Hakala, et al., 2005), which is expected to result in a higher affordance, and hence, in a higher preference.

5 Pilot experiment

To answer the hypotheses stated in chapter 4, it was necessary to do a pilot experiment first, in order to find out which attributes users of mobile devices use to assess a graphical user interface of a mobile device. The subjects were asked to score images representing a graphical user interface on a scale from 1 to 5 for three different attributes. Two of these attributes, affordance and aesthetics, were chosen from the literature. Because assessing a graphical user interface is very subjective, preference was added as a third attribute to the experiment. For the attribute preference, the participant had to assess how much he or she liked the representation of the graphical user interface. To find out which attribute was most appropriate, various variables were used in this experiment, namely, different depth modes, different backgrounds, and different depth cues. The depth mode indicated how the icons and background were positioned with respect to the display screen. The depth cues used were shadowing, luminance differences, and size differences. The variables and values are elaborated in section 5.1, where the stimuli are described. In section 5.2 the equipment setup is discussed, after which in section 5.3 the subjects are discussed. In section 5.4 the protocol is given, followed by the results in section 5.5 and a discussion in 5.6.

5.1 Stimuli

For this experiment the graphical user interface of a mobile device was represented by a background image with five icons superimposed on it. In addition, the graphical user interface contained one icon that was separated from the other five icons. This icon simulated a 'selected icon' on a mobile device. Figure 5.1 shows the background image and the 'selected icon'.



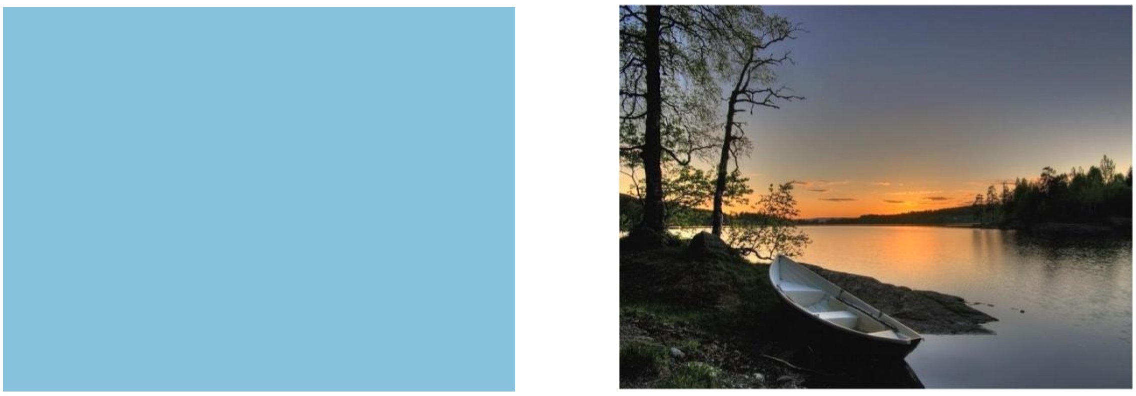
(a) (b)

Figure 5.1 A stimulus consists of (a) a background image, in this case the uniform blue background, with five icons and (b) the 'selected icon' that is rendered in the middle of the bottom row of the background image

Based on this single graphical user interface representation various stimuli were made, written in the 2D + depth format, which was explained in section 2.5. Only the two top quadrants of this format were used. In other words, the stimulus of this experiment did not contain occlusion information.

The stimuli were generated by changing different variables. The variables used were the depth mode, the background content, and some (monocular) depth cues, used nowadays to simulate 3D. Each of these variables was expected to have an effect on the perceived depth in the graphical user interface, and on the assessment attributes.

The content of the background was expected to have an effect on the perceived depth. For example, if the image itself contained perspective, this was expected to increase the perceived amount of depth. To test whether this was true, two different backgrounds were used: the first one was a uniform colored background and the second one a natural scene. These two backgrounds are shown in Figure 5.2.



(a) (b)

Figure 5.2 The two backgrounds used in this experiment (a) the uniform colored background, (b) the natural scene background

When using a 3D display, it is possible to place the background and the 'selected icon' on different depth layers. Additionally, it is also possible to shift the position of these depth layers with respect to the display screen. The specification of the position of the 'selected icon' and background with respect to the display screen was called the depth mode in this experiment. Various depth modes were used, namely, 'sinking background', 'floating icon', and 'on display'. These modes are depicted in Figure 5.3. 'Sinking background' referred to a mode in which the background was behind the display screen (with a grey level of 64 in the depth map of the 2D + depth format), and the 'selected icon' was on the display screen (with a grey level of 128 in the depth map). The mode 'floating icon' referred to a representation, where the background was on the display screen (with a grey level of 128 in the depth map) and the 'selected icon' was rendered in front of the display screen (with a grey level of 192 in the depth map). The 'on display' depth mode rendered both the background and the 'selected icon' on the display screen (with a grey level of 128 in the depth map). As a consequence, there was no disparity (2D) in this depth mode.

(a) (b) (c)

Figure 5.3 Depiction of the various depth modes: (a) Sinking background, with the background behind the display screen and the 'selected icon' on the display screen (b) Floating icon, with the background on the display screen and the 'selected icon' in front of the display screen, and (c) On display, with both the background and 'selected icon' on the display screen.

The standard size of the icons was set to 64 x 64 pixels. This size was neither too big nor too small in the overall view of the graphical user interface. This size was always kept constant for the icons on the background – the not selected icons. The 'selected icon' could have two different sizes: either the standard size or a bigger size, being 95 x 95 pixels.

The variable shadow referred to whether there was a drop-shadow added to the 'selected icon' or not, which is illustrated in Figure 5.4a and b. The shadow was made using Photoshop, and had a size of a quarter to the right and a quarter downwards based on creating a natural appearance. The variable luminance was changed by brightening the 'selected icon'. The icon either had its standard luminance level or the luminance was increased by 70%, which was simulated by using the brightening function in Photoshop. This is illustrated in Figure 5.4c and d.

(a) (b) (c) (d)

Figure 5.4 Variants of the 'selected icon': (a) normal icon, (b) icon with drop-shadow, (c) a 70% brighter icon, and (d) a 70% brighter icon with drop-shadow

The whole set of stimuli consisted of combinations of all values of these variables, which are listed in Table 5.1, except for the depth mode 'on display'. For this depth mode only the background content was varied. As a result, the total amount of stimuli was: 2 depth modes x 2 backgrounds x 2 shadows x 2 luminance levels x 2 sizes + 'on display' x 2 backgrounds = 34. The whole list of combinations can be found in Appendix A, Table 1.

Table 5.1 Variables used in the pilot experiment and their respective values

Variable	Values
Depth mode	Sinking background Floating icon On display
Background	Uniform Natural
Shadow	Shadow No shadow
Luminance	Extra luminance Normal luminance
Size	Big size Normal size

5.2 Equipment setup

The target 3D display (for the mobile device) was a VGA display of 4.5" diagonal with a spatial resolution of 640 x 480 pixels and an optimal viewing distance of 40 cm. The pilot experiment, however, had to be done on a Philips 19" auto-stereoscopic display. The spatial resolution of this display was 1600 x 1200 pixels with a pixel pitch (size of one pixel) of 0.255 mm, and it had an optimal viewing distance of 40 cm. To be able to compare the display used to the target display, not the whole screen area of the 19" display was used. The stimuli had a resolution of 359 x 269 pixels, which resulted in an image size equal to the 4.5" diagonal of the target display. With only 4.5" used the optimal viewing distance was kept to 40 cm.

The Philips 19” display was an auto-stereoscopic 9-view display, with a viewing cone of 34 degrees, which means that the display had 34/9 = 3.8 degrees per view to display the 3D images. This display was only able to display the 9-interleaved-views format, so in order to be able to use the 2D + depth format a printed circuit board (PCB) to convert the 2D + depth format to the 9-interleaved-views format was used. In front of the display a chinrest was placed at 40 cm, being the optimal viewing distance of the display. The chinrest was used to keep the subject at this distance and to restrict head movements, in order to avoid the perception of ghosting and other artifacts like blurring.

For the subjects to be able to give their scores, scoring forms were placed in front of them between the display and chinrest. The scoring form existed of 34 scoring scales which were continuous, and ranged from 1 to 5, where 1 was the lowest score and 5 the highest. Along the scoring scale every interval of 1 integer was indicated. The subjects could put a cross anywhere on the scale.

5.3 Subjects

To this experiment 12 subjects participated, i.e. nine male and three female. All the subjects were students of the faculty of Electrical Engineering, Mathematics and Computer Science of Delft University of Technology. Their age varied between 21 and 28 years with an average of 23.9 years ($\sigma = 2.4$).

The Randot Stereotest was used to test the stereoscopic vision of the subjects. Subjects needed a minimum of 70 arcsecond in order to be allowed to the experiment. No one was excluded based on this criterion.

5.4 Protocol

As said before the subjects had to do the stereo acuity test – the Randot Stereotest – before participating in the experiment.

Then the subjects had to take place in front of the display with their head in the chinrest. Then they were given three subsequent sets of 34 randomly ordered stimuli. They had to score every stimulus on the scoring form, which they had in front of them. Each set was scored on one attribute. The order of the attributes was the same for all subjects, namely, affordance, aesthetics, and preference. No reference image was present in this experiment, so the scaling was absolute. It took the subjects between 20 to 30 minutes to score all three attributes; therefore, it was possible to do the whole experiment in one session.

The scoring of every attribute was preceded with a short training consisting of 5 stimuli. This training gave the subjects the opportunity to get acquainted with the type of stimuli and with how the variables were used in the stimuli. So, in practice each participant saw three times 5 training stimuli and 34 test stimuli. At the end of the scoring form there was some space for the subject to write down some comments. Especially, for the attribute preference the subject were asked to comment on what they actually scored.

5.5 Results

To find out which of the attributes is most appropriate for assessing a graphical user interface the collected data was analyzed using Analysis of Variance (ANOVA) on the three attributes separately. The criterion for selecting the most appropriate attribute was the number of variables that had a significant effect on the scores of the attribute. Indeed, if more variables have a significant effect on the scores of an attribute, it can be concluded that this attribute is able to distinguish between the various variables of interest.

The ANOVAs were performed with the scores as the dependent variable and including the main effects and two-way interactions of the independent variables. The independent fixed variables were: *depth mode*, *background*, *shadow*, *luminance*, and *size*. The subjects (*name*) were included as an independent random variable, and interactions with the variable *name* were excluded from the analysis. In the following tables only the results of the main effects and the significant two-way interactions of the ANOVA results are reported. For the full ANOVA tables see Table 2, Table 3, and Table 4 in Appendix A.

For the attribute *affordance* the results are shown in Table 5.2. It can be seen that the main effect of *depth mode* ($F(2,379) = 46.298$, $p < 0.001$), *shadow* ($F(1,379) = 3.982$, $p = 0.047$), *size* ($F(1,379) = 121.437$, $p < 0.001$), and *name* ($F(11,379) = 34.320$, $p < 0.001$) on the *affordance scores* is significant. The main effect of *background* and *luminance* is not significant, but the interaction between *background* and *shadow* ($F(1,379) = 7.804$, $p = 0.005$) is. All the other interactions between the variables are found to have no significant effect on the *scores*.

The significant effect of the interaction between *background* and *shadow* results from the fact that the shadow is less visible in the natural background than in the uniform background. When looking at the background in Figure 5.2b, it is clear that this background is very dark at the bottom, where the 'selected icon' with its shadow is positioned.

Table 5.2 ANOVA table for the attribute *affordance*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Depth mode	36.351	2	18.175	46.298	.000
Background	.081	1	.081	.207	.649
Shadow	1.563	1	1.563	3.982	.047
Luminance	.013	1	.013	.032	.858
Size	47.672	1	47.672	121.437	.000
Name	148.203	11	13.473	34.320	.000
Background * Shadow	3.064	1	3.064	7.804	.005

For the attribute *aesthetics* the results are given in Table 5.3, and show that the main effects of *depth mode* ($F(2,379) = 43.088$, $p < 0.001$), *shadow* ($F(1,379) = 10.242$, $p = 0.001$), *size* ($F(1,379) = 15.787$, $p < 0.001$), and *name* ($F(11,379) = 20.934$, $p < 0.001$) on the *scores* is significant. Again, the variables *background* and *luminance* have no significant effect on the *scores*. Also for this attribute the interaction between *background* and *shadow* ($F(1,379) = 6.089$, $p = 0.014$) is significant.

Table 5.3 ANOVA table for the attribute *aesthetics*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Depth mode	34.621	2	17.311	43.088	.000
Background	.708	1	.708	1.761	.185
Shadow	4.115	1	4.115	10.242	.001
Luminance	.195	1	.195	.485	.487
Size	6.342	1	6.342	15.787	.000
Name	92.510	11	8.410	20.934	.000
Background * Shadow	2.446	1	2.446	6.089	.014

The results for the attribute *preference* are given in Table 5.4, and show that none of the interactions between the variables have a significant effect on the *scores*. The main effect of *depth mode* ($F(2,379) =$

70.369, $p < 0.001$), *size* ($F(1,379) = 62.925$, $p < 0.001$), and *name* ($F(11,379) = 6.392$, $p < 0.001$) on the *scores* is significant. The variables background, *shadow*, and *luminance* have no significant effect on the *preference scores*.

Table 5.4 ANOVA table for the attribute *preference*

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Depth mode	70.284	2	35.142	70.369	.000
Background	1.055	1	1.055	2.113	.147
Shadow	.160	1	.160	.321	.571
Luminance	.430	1	.430	.861	.354
Size	31.425	1	31.425	62.925	.000
Name	35.111	11	3.192	6.392	.000

From the ANOVA results it can be concluded that the attributes *affordance* and *aesthetics* are affected by the same significant variables. This might suggest that both attributes represent the same assessment criterion. To check this hypothesis, the trend in the variables for both attributes is evaluated by comparing the variable value with the highest mean score for each attribute. These values with their mean score and standard deviation are given in Table 5.5 and Table 5.6 for the attributes *affordance* and *aesthetics*, respectively. Comparing the two tables shows that both attributes do not have the same trend and so, do not represent the same assessment criterion.

Table 5.5 Highest mean scores per variable for *affordance*

Variable	Value	Mean (μ)	Std. dev (σ)
Depth mode	Sinking background	3.71	0.84
Background	Natural background	3.45	1.00
Shadow	Shadow	3.55	0.89
Luminance	Luminance	3.49	0.92
Size	Big size	3.84	0.81

Table 5.6 Highest mean scores per variable for *aesthetics*

Variable	Value	Mean (μ)	Std. dev. (σ)
Depth mode	On display	3.51	0.95
Background	Natural background	2.97	0.83
Shadow	No shadow	3.00	0.87
Luminance	No luminance	2.93	0.88
Size	Big size	2.96	0.85

A similar comparison can be made between the attributes *affordance* and *aesthetics*, on the one hand, and the attribute *preference* on the other hand. To do so, the variable values with the highest mean score in *preference* are given in Table 5.7 together with the mean score and standard deviation. Comparing Table 5.7 to Table 5.5 and Table 5.6 demonstrates that especially the attributes *aesthetics* and *preference* have almost the same trend.

Table 5.7 Highest mean scores per variable for *preference*

Variable	Value	Mean (μ)	Std. dev. (σ)
Depth mode	Sinking background	3.46	0.81
Background	Natural background	3.11	0.91
Shadow	No shadow	3.03	0.94
Luminance	No Luminance	3.05	0.93
Size	Big size	3.32	0.84

A similar conclusion can be drawn from a Pearson correlation analysis, calculating the mutual correlation between the three attributes. This analysis shows that the attributes *affordance* and *aesthetics* have a very low correlation, $r = 0.237$, $N = 408$, $p < 0.001$. The attributes *affordance* and *preference* have a slightly higher correlation, $r = 0.343$, $N = 408$, $p < 0.001$, whereas the attributes *aesthetics* and *preference* have the highest correlation, $r = 0.526$, $N = 408$, $p < 0.001$. Nonetheless, even the highest correlation is still moderate.

Based on the comments given by the subjects at the end of the experiment it can be concluded that the size of the shadow was too large, the disparity difference between 'selected icon' and background too high, and the luminance increase for the 'selected icon' too small. Subjects commented that the difference in luminance was hardly noticeable and that the stimuli became slightly blurred.

5.6 Discussion

Based on our results, it is clear that preference is not the best attribute to use in the assessment of a graphical user interface of a mobile device. The correlation between the preference and aesthetics scores is moderate, while the trend of both attributes with respect to the variables used in the experiment is almost the same. The attributes affordance and aesthetics clearly represent different assessment criteria. They have a low mutual correlation and their trend in relation to the variations in the variables is not the same. This implies that both attributes are complementary, and therefore, may be useful for the assessment of a graphical user interface of a mobile device.

The results of this pilot experiment already reveal some interesting aspects related to the variation in the variables. For the attributes affordance and preference the depth mode 'sinking background' has the highest mean. Also previous research, where the performance of reading was measured with different depth modes, showed that the 'sinking background' mode gave a better performance than the 'floating text' mode (Mizobuchi, et al., 2008). Hence, our results are in line with the literature, suggesting that a 'sinking background' mode is more appropriate than a 'floating text (or icon)' mode from the point of view of reading performance, affordance and preference. The reason may be that subjects mainly look at the foreground object (text or icon) and do not mind that the background becomes slightly blurry, when rendered at a higher depth value, as long as the foreground object remains sharp.

For the attributes aesthetics and preference the value 'no shadow' has the highest mean, which implies that the subjects prefer to have no shadow added to the icons. They do prefer the big sized icon above the standard sized icon, but they do not prefer the addition of luminance to the selected icon.

As mentioned above, the values for the size of the shadow and the luminance increase were chosen based on resulting in a natural appearance. This natural appearance was judged by a limited number of people, which, of course, may have been insufficiently representative. Therefore, it is possible that the values for the size of the shadow and the percentage of luminance increase were too large or too small, as is also suggested by the comments given by the subjects at the end of the experiment. Hence, to find the best values for these variables, another experiment is needed.

6 Tuning experiment

Before being able to do the main experiment the questions that have arisen from the results of the pilot experiment have to be answered. In order to answer these questions a tuning experiment is performed, investigating the preference of subjects with respect to a number of settings. The related questions to be answered are:

- *What is the preferred disparity for different depth modes and backgrounds?*
  It is possible that for different depth modes and backgrounds the preference in disparity is different.
- *What is the preferred size of the 'selected icon' for different depth modes and disparities?*
  If two of the same objects are placed at a distance from each other, the object closer to the viewer looks bigger than the one that is further away from the viewer. This is called the linear perspective monocular depth cue. When an icon that comes out of the screen is larger than the icons in the background, this might look more natural due to this linear perspective depth cue. To investigate whether this is the case with icons in a user interface, different sizes for the icons are evaluated for various depth modes and disparities.
- *What is the preferred luminance of the 'selected icon' for different backgrounds and disparities?*
  Differences in brightness are known to create a kind of 'pop-out' effect (achromatic pop-out) (Theeuwes & Lucassen, 1993), and with that may enhance the feeling of depth. This 'pop-out' effect can be implemented in a user interface by giving the 'selected icon' some higher luminance than the icons on the background, but it is not known yet how big that luminance increase should be. Since the perceived luminance increase is expected to depend on the contrast ratio between the 'selected icon' and background, two different backgrounds are used.

The stimuli used for the tuning experiment are elaborated in section 6.1. The tuning experiment is done with different subjects and equipment than the pilot experiment (described in chapter 5). The new conditions are discussed in sections 6.2 and 6.3, respectively. The protocol of this experiment is given in section 6.4, while the results are shown in section 6.5, followed by a discussion in section 6.6.

6.1 Stimuli

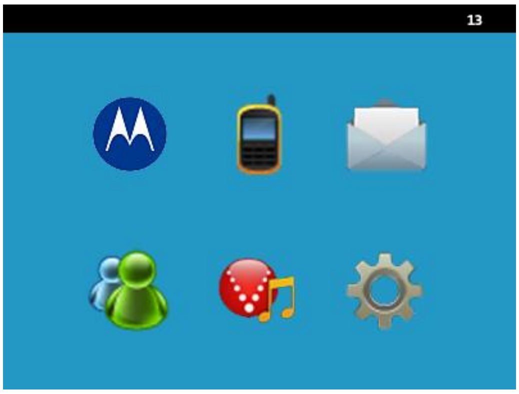


Figure 6.1 Stimulus used in the tuning experiment with a background containing five 'non-selected' icons and the 'selected icon' in the middle of the bottom row.

Since each of the questions mentioned above was investigated separately, different sets of stimuli had to be created for each part of the tuning experiment. When the disparity or size was tuned, the stimuli existed of a 'selected icon' on a background with five 'non-selected' icons, where the 'selected icon' was positioned in the middle of the bottom row. The icons used were the same as shown in chapter 5. The 'selected icon' received a different size, and/or disparity, depending on the specific research question. Also the background varied depending on the specific research question. Figure 6.1 shows an example of a typical stimulus. When the luminance was tuned, the background contained six equal icons positioned at different disparities, which will be explained in more detail in section 6.1.3.

The stimuli for all tunings contained a black ribbon with a number on top of the image, which was not part of the graphical user interface. The number referred to a particular stimulus in the set, but was attributed randomly to the stimuli to avoid the possibility that subjects could pick systematically the same disparity, luminance or size for the different depth modes and/or backgrounds. The number was just there for the experimenter to be able to write down the tuned image.

The icons used were only available in portable network graphics (png) format, so when they had to be scaled to a larger size, the icons became slightly blurred. The sizes of the icons were chosen in such a way as to keep the amount of blurring as small as possible. This was done by making the icons in different sizes and looking whether they were still sharp enough to be used for the experiment. The icons also had a standard shadow. To be able to do the experiment, the shadow had to be removed for all icons. This was accomplished by using Photoshop. However, it can be seen in Figure 6.2(a) that the icon without shadow is slightly jagged compared to the icon with shadow in Figure 6.2(b). Vector images would have been ideal to do these operations on, but unfortunately the vector images of these icons were not available.

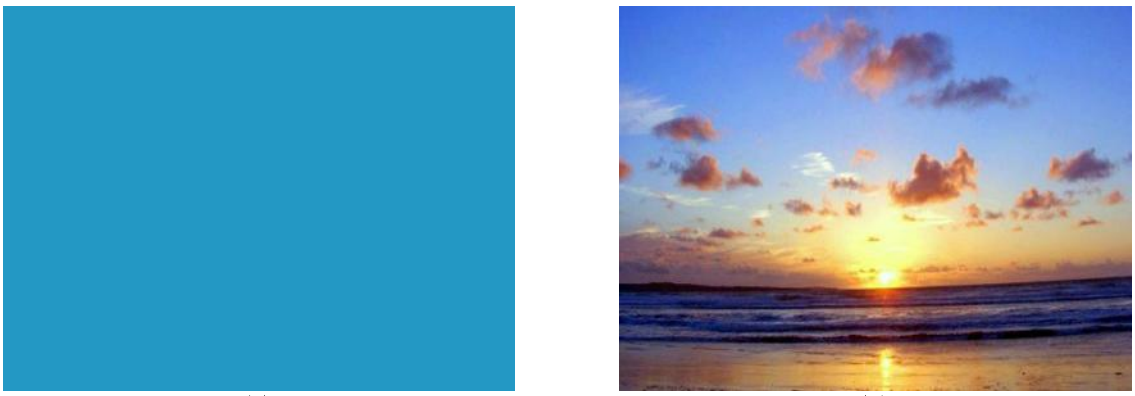
(a) (b)

Figure 6.2 The icon that was used as the 'selected icon' (a) without shadow (b) with shadow

The stimuli were composed with the 2D + depth format, containing four quadrants. Each quadrant had a resolution of 400 x 300 pixels. For this experiment the third and fourth quadrant were not used.

6.1.1 Tuning of disparity

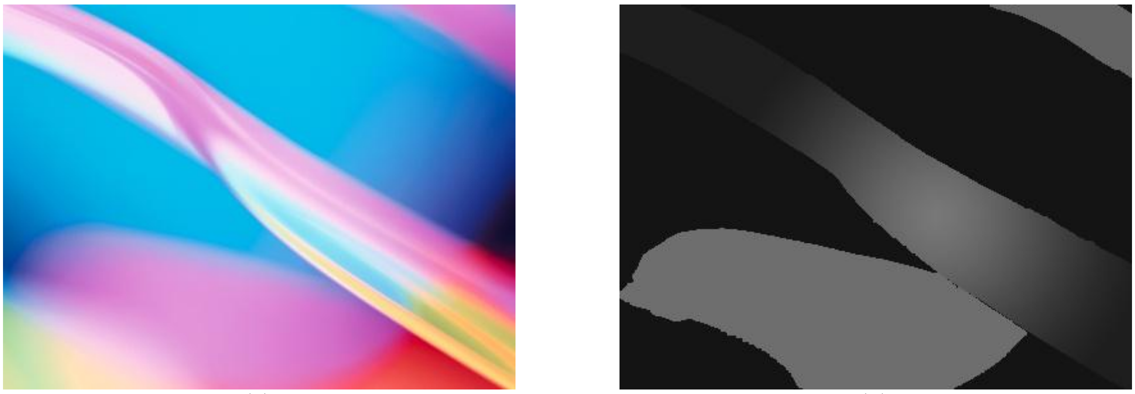
In the first experimental part, the disparity of either the foreground or background was tuned, depending on the depth mode. Three background images were used: two backgrounds without disparity and one background which already had disparity. The uniform background (Figure 6.3(a)) and natural background (Figure 6.3(b)) had no disparity.



(a) (b)

Figure 6.3 The backgrounds used in the tuning of disparity: (a) the uniform blue background, and (b) the natural background

The image of the background with depth is given in Figure 6.4(a) with its depth map shown in Figure 6.4(b).



(a) (b)

Figure 6.4 The background image with depth used in the tuning of disparity: (a) image of the background, and (b) the depth map of the background

Three depth modes were used: 'sinking background' and 'floating icon', and 'window effect'. The depth modes 'sinking background' and 'floating icon' were implemented in the same way as in the pilot experiment, described in chapter 5. The depth mode 'window effect' was characterized by the background being always behind the display screen and the 'selected icon' being either on the display screen or also behind it. This depth mode created the impression of looking through a window; hence, its name.

For each depth mode 10 disparity levels were selected, as given in Table 6.1 by the corresponding grey level in the depth map for both the background and the 'selected icon'. It should be mentioned that the 'selected icon' in the case of the background with depth had the same disparity value as the 'selected icon' in the 'floating icon' depth mode.

Table 6.1 Disparity in grey level for the depth modes: sinking background, floating icon and window effect. The background with depth used the same grey levels for the 'selected icon' as the 'selected icon' in the floating icon depth mode.

Sinking background		Floating icon		Window effect	
Background (in grey level)	**'Selected icon' (in grey level)**	**Background (in grey level)**	**'Selected icon' (in grey level)**	**Background (in grey level)**	**'Selected icon' (in grey level)**
20	128	128	128	20	128
30	128	128	140	30	120
35	128	128	150	40	110
40	128	128	160	40	100
50	128	128	170	50	100
70	128	128	180	50	90
80	128	128	185	60	90
100	128	128	190	60	80
110	128	128	195	70	80
128	128	128	210	75	75

The number of tunings in disparity that the subjects had to perform was: 2 backgrounds (uniform and natural) x 3 depth modes + 1 background (with depth) = 7.

6.1.2 Tuning of Size

The basic form of the stimulus for tuning the size of the 'selected icon' was the same as that for tuning the disparity. Only the size of the 'selected icon' was varied. Table 6.2 shows the eleven sizes (in pixel counts) that were used.

Table 6.2 The sizes used for the 'selected icon' in the tuning

Size (in pixels)
60 x 60
66 x 66
72 x 72
78 x 78
84 x 84
90 x 90
96 x 96
102 x 102
108 x 108
114 x 114
120 x 120

For this tuning only the uniform background, two depth modes and two disparities were used. The 'sinking background' mode was used with disparities in grey level of 64 and 116 for the background, while the icon remained at the screen (i.e. at a grey level of 128). The 'floating icon' mode was used with disparities in grey level of 140 and 192 for the icon, while the background remained at the screen.

The total number of tunings in size that the subjects had to do was: 1 background x 2 depth modes x 2 disparities = 4 tunings with each 11 levels in size.

6.1.3 Tuning of luminance

The stimuli for tuning the luminance were different than for the other tunings. Instead of different icons, six equal icons were used, as shown in Figure 6.5.

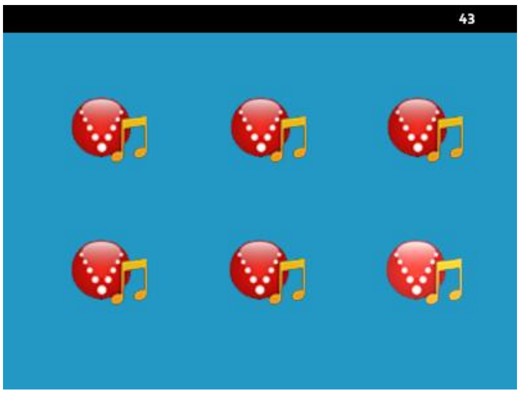


Figure 6.5 The stimulus with the uniform background and the six equal icons, used in the tunings of the luminance

The top three icons were considered as reference, and were not changed. The three icons at the bottom changed one by one in luminance. A change in luminance was simulated by adjusting the brightness of the icon in Photoshop. The brightness increased in steps of 10% to 80% of the original brightness.

Each column icons was put at a different disparity. The leftmost column was put 'on display' with a grey level of 128, the middle column was put at 'half disparity', and the rightmost column was put at 'maximal disparity'. It should be noted that the maximal disparity in this case did not correspond to a grey level of 255, but to a grey level of 180 in order to limit the amount of blur on the icon. As a consequence, the 'half disparity' corresponded to a grey level of 154.

Two uniform colored backgrounds were used, as shown in Figure 6.6. The choice of a light blue and a dark blue background was made, because the preferred luminance was expected to depend on the contrast ratio between icon and background. The light blue background was generated with a RGB-value of (35, 152, 196), and the dark blue background was created with a RGB-value of (11, 47, 60).

(a) (b)

Figure 6.6 The backgrounds used in the tunings of luminance: (a) the light blue background, and (b) the dark blue background

For this part of the experiment a total number of 2 backgrounds x 3 disparities = 6 tunings had to be done. Each tuning contained 9 different luminance levels.

6.2 Equipment setup

An 8.4" Philips auto-stereoscopic display with a resolution of 800 x 600 was used. This display had an optimal viewing distance of 1 meter. On this display a 4.5" display with a resolution of 480x360 and an optimal distance of 40 cm was simulated. This means that the two displays differed in all variables by about a factor of 2. As a consequence, the 4.5" display could be simulated on the 8.4" display, by displaying the stimuli at a resolution of 800x600, and increasing the viewing distance to 1 meter. In that way, the angular resolution and the angular image size of the 8.4" display corresponded to what was intended for the 4.5" display. The remote control of the display was used to operate the display.

A chinrest was placed at one meter in front of the display, to keep the subjects at a fixed distance.

During each tuning the image preferred by the subject was filled in in an excel file at a laptop that was used during the experiment.

6.3 Subjects

To this experiment 37 subjects participated. 18 of which were employees of the Motorola, Inc lab. Their average age was 45 years ($\sigma = 9.3$). The other 19 subjects were Master and PhD students of Arizona State University. Their average age was 25 years ($\sigma = 3.2$). The average age of all subjects was 35 years ($\sigma = 11.9$).

Each subject had to perform a Randot Stereotest prior to the experiment, and was allowed to participate with a stereo vision capability of 70 arcsecond or lower. This is a relatively low threshold, with the consequence that none of the subjects was excluded from the tuning experiment.

6.4 Protocol

Before starting the experiment each subject got a short verbal introduction to the experiment, wherein the goal of the experiment and what was expected from the subject was explained. Then they had to do the Randot Stereotest. After this test they were seated in front of the display with their head in the chinrest.

Every subject performed the tunings in the same order, namely he or she started with the tunings of the disparity, then performed the tunings of the icon size, and finally performed the tunings of the luminance. For each tuning the subject was allowed to step back and forth between the various images in the stimulus set, and was asked to pick the image with the preferred disparity, size or luminance depending on the tuning. A black image indicated the beginning or the end of a set of stimuli belonging to a given tuning. The subject did not operate the remote control of the display himself, since the display reaction time of the remote control was very long. Hence, the subject could become impatient and press the button again. When this was done the display skipped an image and displayed the next image. This had to be avoided in this experiment. Therefore the experimenter operated the remote control. When the subject saw the preferred stimulus, its number was typed into the excel file by the experimenter and the next tuning was started.

The whole experiment, including all 7 + 4 + 6 = 17 tunings took about 45 minutes.

6.5 Results

Disparity

Before analysis, the results of the preferred disparity were normalized by calculating the difference in grey level value between the icon and background. Then, an ANOVA was performed with the normalized *disparity* as dependent variable, the *background* and *depth mode* as fixed factors, and the *outcome of the Randot Stereotest* as covariate, including the two-way interaction between *background* and *depth mode*. The *outcome of the Randot Stereotest* had no significant effect on the preferred *disparity* ($F(1,251) = 0.779$, $p = 0.378$). The *depth mode* ($F(2,251) = 36.425$, $p < 0.001$) and *background* ($F(1,251) = 10.860$, $p = 0.001$) had a significant effect on the *disparity*.

Since based on the results of this first ANOVA we could conclude that the variability in preferred disparity among participants did not result from differences in stereoscopic vision ability, we repeated the ANOVA, but now with the subjects' *name* as a random factor. Again the normalized *disparity* was the dependent variable, the *background* and *depth mode* were included as fixed factors, including also their two-way interaction. The main effect of *depth mode* ($F(2,216) = 50.443$, $p < 0.001$), *background* ($F(1,216) = 15.040$, $p < 0.001$) and *name* ($F(36,216) = 3.685$, $p < 0.001$) on the preferred disparity was significant. No significant interaction effect between *background* and *depth mode* ($F(2,216) = 1.208$, $p = 0.301$) was found.

In Table 6.3 the mean value and standard deviation in absolute disparity of the preferred *disparity* for each *depth mode* and *background* is given in grey levels.

Table 6.3 Mean value and standard deviation of the preferred *disparity* in grey levels for each *depth mode* and *background*

Depth mode	Background	Mean(μ)	Std. dev (σ)
Sinking background	Uniform	43	19
	Natural	57	25
Floating icon	Uniform	181	14
	Natural	175	11
Window effect	Uniform	64	24
	Natural	58	25
Non	Depth	177	14

Size

To perform an ANOVA on the collected data of the preferred *size*, we first normalized the disparity levels, by calculating the difference in grey level between the display screen (with a grey level of 128) and the icon or background (with a grey level of 64, 116, 140, and 192 depending on the depth mode).

The ANOVA was performed with the preferred *size* as dependent variable, the *depth mode* and normalized *disparity* as fixed factors and the subjects' name as random factor. Also the two-way interaction between *depth mode* and normalized *disparity* was added. Only the variable *name* ($F(36,108) = 3,594$, $p < 0.001$) had a significant main effect on the preferred *size*. The *depth mode* ($F(1,108) = 0.219$, $p = 0.640$) and normalized *disparity* ($F(1,108) = 2.246$, $p = 0.137$) did not significantly affect the preferred *size*, neither did the interaction between *depth mode* and normalized *disparity* ($F(1.108) = 0.562$, $p = 0.455$).

To know the preferred *size*, its mean and standard deviation was calculated for the two depth modes (see Table 6.4) and the two disparity ranges (see Table 6.5). Indeed, all mean values were nearly equal, at least with small variations well within the standard deviation.

Table 6.4 The mean of the preferred *size* for the different *depth modes*

Depthmode	Mean	N	Std. Deviation
Floating Icon	133.78	74	21.499
Sinking Background	135.14	74	23.481
Total	134.46	148	22.445

Table 6.5 The mean of the preferred *size* for the different *disparities*

Disparity	Mean	N	Std. Deviation
12	136.62	74	23.307
64	132.30	74	21.489
Total	134.46	148	22.445

So, changing the *depth mode* or the *disparity* level had no influence on the preferred *size*. The preferred *size* was around 134% of the normal size. A box plot, shown in Figure 6.7, confirmed that the mean preferred *size* for all disparities was around 134% of the original size, with, however, a huge spread due to individual differences in preferred *size*.

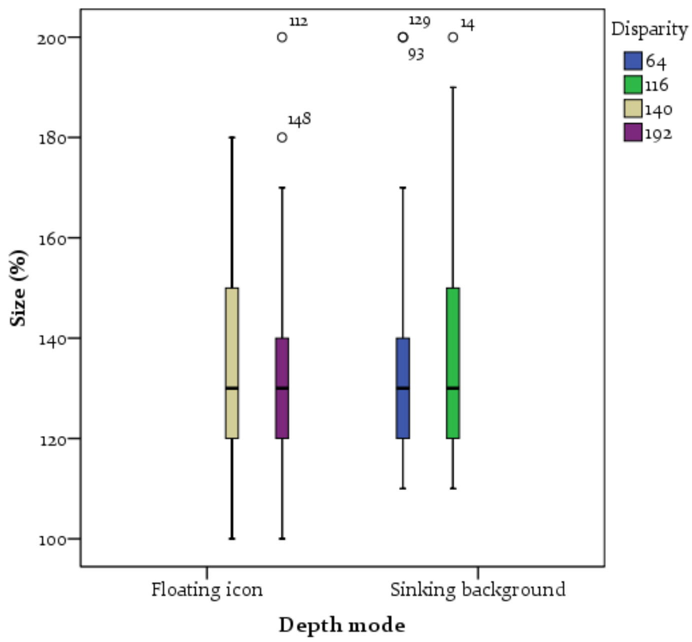


Figure 6.7 Box plot on the not normalized data of the preferred *size*

Luminance

An ANOVA was performed with the preferred *luminance* as the dependent variable, the *background* and *disparity* as the fixed factors, including their two-way interaction and the subjects' *name* as random factor. No significant effect on the preferred *luminance* was found for the *background* ($F(1,180) = 0.019$, $p = 0.892$), the *disparity* ($F(2,180) = 0.002$, $p = 0.998$), and their interaction ($F(2,180) = 0.488$, $p = 0.615$). The subjects' *name* ($F(36,180) = 5.828$, $p < 0.001$) did have a significant main effect on the preferred *luminance*.

A box plot of the raw data, given in Figure 6.8, shows that all subjects stay within the same range of luminance for every disparity level. The preferred *luminance* level is about 30% on top of the original luminance level. The spread among subjects, however, is very large, almost covering the full range of luminance levels used in the experiment.

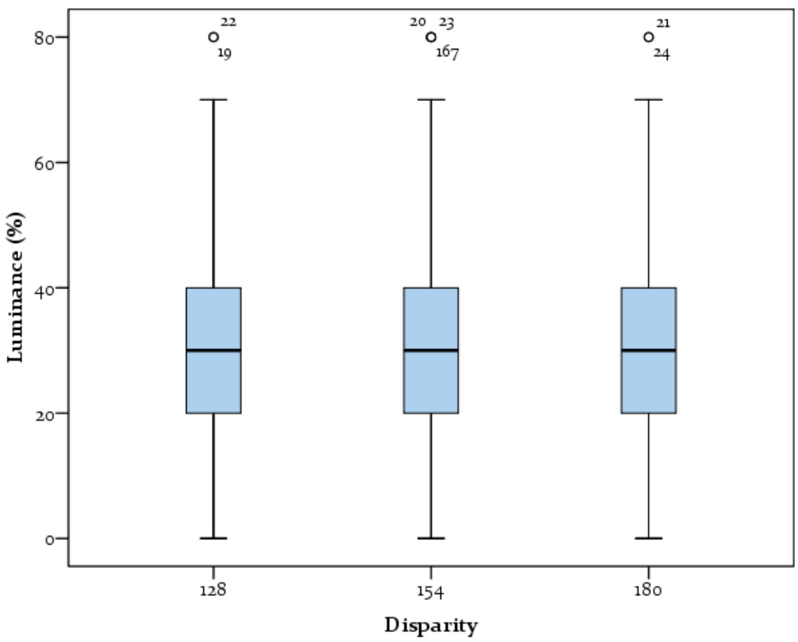


Figure 6.8 Box plot of the preferred *luminance*

6.6 Discussion

This experiment was conducted to find the preferred disparity, size of the 'selected icon', and luminance of the 'selected icon'. The results from this experiment were needed to determine the values of these variables to be used in the main experiment.

One problem with the results of the tuning experiment was the big spread among the subjects in preferred disparity, size and luminance. For each type of tunings the subjects' name had a significant effect on the preference. In other words, giving in the main experiment the disparity, size and luminance of the 'selected icon' the mean preferred value of the tuning experiment would be too big for some subjects, and too small for other subjects. Because of this observation, it would be ideal to give each participant of the main experiment, his or her personal preferred settings for the disparity, size and luminance of the 'selected icon'. But, in that case, it might become difficult to compare the results of the main experiment among subjects, and to draw general conclusions.

An option to reduce the problem of spread among subjects would be to select a smaller group of (e.g. 20 out of the 37) subjects with a more uniform preference.

To select such a group of subjects with more or less the same preference in disparity a post-hoc test was performed on the tuning results of the disparity. This resulted in one subset of 29 subjects who had a non-significantly different preference in disparity. From this subset 20 subjects were asked to participate again in the main experiment. For these 20 subjects the average preference in disparity for the various depth modes and backgrounds are given Table 6.6. Additionally, their mean preference in size was 136 with a standard deviation of 22 and their mean preference in luminance was 29 with a standard deviation of 18. The mean preference of size and luminance of these subjects was about the same as the overall mean preference of these two variables.

Table 6.6 The mean value and standard deviation of the disparity for the various depth modes and backgrounds for the subjects who will participate in the main experiment

Depth mode	Background	Mean(μ)	Std. dev (σ)
Sinking background	Uniform	40	15
	Natural	53	24
Floating icon	Uniform	181	13
	Natural	176	12
Window effect	Uniform	69	24
	Natural	62	20
None	Depth	178	14

The preferred disparity also depended on the depth mode and the background. This dependency will be taken into account in the definition of the disparity settings for the main experiment. The preferred size of the 'selected icon' was independent of the disparity of the icon and of the background. In all cases people preferred on average an increase in size of 134% for the 'selected icon'. In other words, people preferred the 'selected icon' to be somewhat bigger than the original icon, independent on whether the icon was on the display screen or floating in front of it.

The results of the tuning in luminance showed that subjects preferred an increase in the luminance of about 30% for the 'selected icon' with respect to the 'non-selected icons'. This preferred increase was independent of the disparity level. Despite of these results, we decided to leave the variable luminance out of the main experiment. An increase of 30% in luminance is very hard to detect without any reference material. Also various subjects said that they didn't mind whether the luminance was increased or not.

7 Main experiment

In the main experiment the main research question is answered. This question is:

"When using an auto-stereoscopic display on a mobile phone, which depth cues can be used in the user interface to increase its overall experience?"

To answer this question, subjects were requested to assess four attributes comprising the overall experience: namely affordance, aesthetics, image quality and perceived amount of depth. Affordance and aesthetics already showed their added value in the pilot experiment, described in chapter 5. Image quality and perceived amount of depth were added, since it is known from literature that these two attributes affect the viewer's preference for 3D stimuli (Seuntiëns, et al., 2005).The attributes were assessed on stimuli of graphical user interfaces built along the guidelines found in the tuning experiment, described in chapter 6.

First, section 7.1 explains how the subjects were selected. Then the stimuli used in this experiment are discussed in section 7.2, while the experimental setup is given in section 7.3. The protocol of this experiment is discussed in section 7.4. The results of the experiment are given in section 7.5 and discussed in section 7.6.

7.1 Subjects

In order to avoid that our results would be confined by a large difference in preferred disparity among the participants, we decided to select for the main experiment those twenty people that tuned their preferred disparity level most comparably in the tuning experiment, as already discussed in section 6.6. These twenty subjects were selected applying a Tukey post-hoc test to the ANOVA results of the tuned disparities. So, the analysis was performed with the preferred *disparity* as the dependent variable, and the subjects' *name*, *depth mode*, and *background* as fixed factors. The subjects' *name* was used as a fixed factor in this analysis in order to be able to apply the post-hoc test to this variable. The post-hoc test was used in order to see how the group of subjects was clustered with respect to their preferred disparity. The full table of the results is given in Appendix C, Table 8. It shows that the second and fifth subset of subjects had the highest significance value, indicating that the difference in preferred disparity among the subjects was smallest for these subsets. In the second subset the difference between the lowest and the highest mean per subject was larger than in the fifth subset. Therefore, the fifth subset was selected, and consequently, subjects 11 till 30 were asked to participate again in the main experiment.

By coincidence half of the subjects were from Motorola, Inc and half from ASU. Their average age was 35 years, with $\sigma = 12$ years.

7.2 Stimuli

The basic layout of the stimuli was the same as for the tuning experiment; an example of such a stimulus was already shown in Figure 6.1. Again, only one of the six icons on the background was the 'selected icon' and only this 'selected icon' was changed in appearance. The rest of the icons was kept constant in their appearance.

The variables used to change the stimuli were: background, depth mode, shadow, and size. As in the tuning experiment two backgrounds without disparity and one background with disparity were used (see Figure 6.3 and Figure 6.4, respectively). Three depth modes were used for the two backgrounds without disparity, namely sinking background, floating icon, and window effect. For the background with disparity none of the depth modes was used. Table 7.1 shows the mean disparity value (and standard deviation) for the different depth modes as tuned during the tuning experiment by these subjects that were selected again for the main experiment. The shadow that was already present in the icon had a size of 6.25% to the right and 7.7% to the bottom of the icon. Additionally, two sizes were used for the 'selected icon'; the original size of 60 x 60 pixels and a bigger size of 80 x 80 pixels.

Table 7.1 The mean value and standard deviation of the disparity for the various depth modes and backgrounds, as found in the tuning experiment for the subjects that participated in this experiment

Depth mode	Background	Mean(μ)	Std. dev (σ)
Sinking background	Uniform	40	15
	Natural	53	24
Floating icon	Uniform	181	13
	Natural	176	12
Window effect	Uniform	69	24
	Natural	62	20
None	Depth	178	14

So, the total number of stimuli used for this experiment was: 2 backgrounds (natural & uniform) x 3 depth modes (sinking background, floating icon & window effect) x 2 shadows (with shadow & without shadow) x 2 sizes (normal size & big size) + 1 background (depth) x 2 shadows x 2 sizes + 3 backgrounds (natural, uniform & depth) x 1 depth mode (on display) = 31 stimuli.

7.3 Experimental setup

The display setup for this experiment was the same as for the tuning experiment (see section 6.2).

The only difference was that now the subjects had scoring forms in front of them. The scoring forms had the same scoring scales as used in the pilot experiment (see section 5.2); i.e. they consisted of a continuous scoring scale ranging from 1 to 5, where 1 corresponds to the lowest score and 5 to the highest score. Along the scoring scale every interval of 1 integer was indicated. The subjects were allowed to put their score anywhere on the scoring scale.

7.4 Protocol

When the subject took place before the display, he or she got a short written introduction on what to do in this experiment and on how to use the scoring forms. In this experiment four attributes had to be scored, namely affordance, aesthetics, image quality, and perceived amount of depth. These attributes were scored in a different random order for each of the subjects. The subjects saw all stimuli one after the other, in a different random order per attribute and subject. They were asked to score each stimulus on how appropriate it was with respect to the attribute to be scored, using the absolute scaling method. To get acquainted with the range of variation in the stimuli, each subject saw first five training stimuli.

Because scoring four attributes in one session would have been too tedious, the experiment was divided into two sessions, where two attributes were scored in each session. Which attributes were scored in each session was random per subject. The second session took place one day after the first session. One session took about 20 to 30 minutes.

7.5 Results

All 20 subjects participated in the first session. One subject, however, did not show up for his second session. Therefore, only the results of the remaining 19 subjects were further analyzed.

On each of the attributes perceived depth, image quality, affordance, and aesthetics an ANOVA was performed incluiding the main effects and all two-way interactions of the variables *depth mode*, *background*, *shadow*, and *size*. The subjects were included as a random variable. Only the variables with a significant main effect on the scores were further evaluated to find out the exact effect of the variable on the scores. The significant interactions were plotted in a box plot to illustrate the effect of the interaction.

In the description of the stimuli it can be noticed that we did not have a full block design. The 'on display' depth mode, for example, only consisted of 3 stimuli, in which only the background was varied. The depth modes 'sinking background', 'floating icon', and 'window effect' each consisted of 8 stimuli, in which the background, shadow, and size were varied. The depth mode 'none' consisted of 4 stimuli; since this depth mode is linked to the background with depth, it only was changed for the shadow and size variables. Because of this experimental design, the data was analyzed in two ways, namely, the data was analyzed for the incomplete design as though all variables were combined to each other and the data was analyzed for the depth modes that had a full block design (i.e. 'sinking background', 'floating icon', and 'window effect').

7.5.1 Perceived depth

Incomplete design analysis

For the attribute *perceived depth*, an ANOVA was done with the score of the *perceived depth* as dependent variable. The *depth mode*, *background*, *shadow*, and *size* were used as fixed factors, including their two-way interaction and the subjects' *name* was added as a random factor. The full overview of the results can be found in Appendix C, Table 9. The *depth mode* ($F(4,550) = 96.947$, $p < 0.001$), *background* ($F(2,550) = 3.342$, $p = 0.036$), *size* of the icon ($F(1,550) = 60.157$, $p < 0.001$) and *name* of the subject ($F(18,550) = 21.849$, $p < 0.001$) all had a significant effect on the *perceived depth* scores. No significant main effect on the *perceived depth* was found for the variable *shadow* ($F(1,550) = 2.424$, $p = 0.120$).

The effects of the *depth mode* on the *perceived depth* score was further evaluated with a Tukey post-hoc test. Four subsets were found as shown in Table 7.2. The perceived amount of depth was lowest for the depth mode 'on display', while it was highest for the 'sinking background' depth mode. The perceived depth for the depth modes 'floating icon' and 'none', where the icon was in front of the display, was significantly lower than for those depth modes where the background was behind the display. The perceived depth for the depth mode 'window effect' was significantly lower than for the depth mode 'sinking background'. From these findings it can be concluded that the perceived depth was highest when the icon had no disparity at all.

Table 7.2 Results of a Tukey post-hoc test on the variable *depth mode* for the *perceived depth* scores

Depth mode	Subset			
	1	2	3	4
On Display	1.41			
Floating Icon		3.15		
None		3.16		
Window Effect			3.56	
Sinking Background				3.86
Sig.	1.000	1.000	1.000	1.000

A Tukey post-hoc test was also performed on the variable *background* to further analyze its effects on the perceived amount of depth. Table 7.3 reveals three subsets, indicating that all backgrounds used were significantly different from each other. The perceived amount of depth was lowest for the background with depth and highest for the natural background.

Table 7.3 Results of a Tukey post-hoc test on the variable *background* for the *perceived depth* scores

Background	Subset		
	1	2	3
Depth	2.85		
Uniform		3.26	
Natural			3.44
Sig.	1.000	1.000	1.000

When the 'selected icon' was larger than the original icon size, the perceived amount of depth was higher. The mean score of the perceived amount of depth was $\mu = 3.70$ with $\sigma = 0.88$ when the 'selected icon' was bigger. For the original icon size the mean score of the perceived depth was $\mu = 2.92$ with $\sigma = 1.11$.

Only for *background* and *shadow* a significant interaction on the *perceived depth* score was found ($F(1,550) = 5.399$, $p = 0.021$). An illustration of that interaction can be found in Figure 7.1. The effect of background on the perceived amount of depth is similar with or without shadow, but is more pronounced in the absence of a shadow.

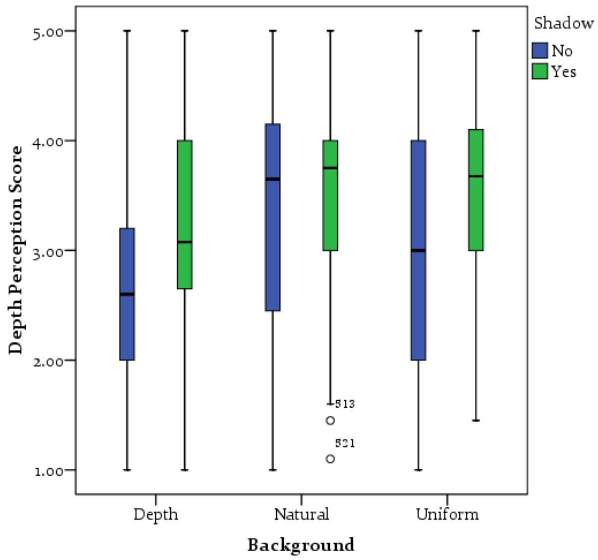


Figure 7.1 Illustration of the interaction between *background* and *shadow* on *perceived depth*

Full block design analysis

For the full block design analysis, an ANOVA of the attribute *perceived depth* was done at the same manner as described above. With the score of the *perceived depth* as dependent variable and the *depth mode*, *background*, *shadow*, and *size* were used as fixed factors, including their two-way interaction and the subjects' *name* was added as a random factor. The full overview of the results can be found in Appendix C, Table 10. The *depth mode* ($F(2,423) = 54.964$, $p < 0.001$), *background* ($F(1,423) = 9.820$, $p = 0.002$), *size* of the icon ($F(1,423) = 66.790$, $p < 0.001$) and *name* of the subject ($F(18,423) = 25.013$, $p < 0.001$) all had a significant effect on the *perceived depth* scores. No significant main effect on the *perceived depth* was found for the variable *shadow* ($F(1,423) = 0.817$, $p = 0.367$).

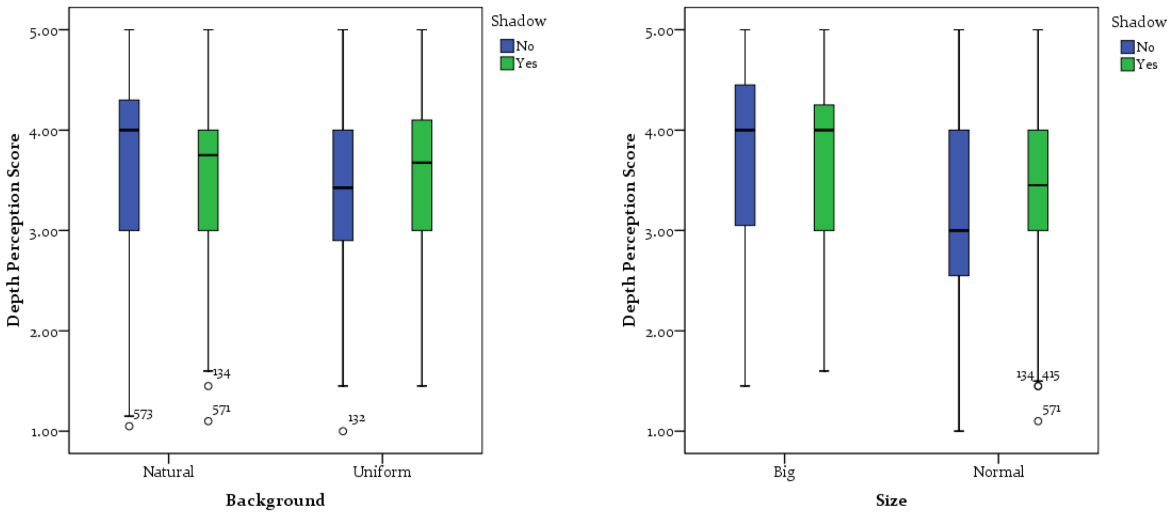
A Tukey post-hoc test was performed to evaluate the effects of the *depth mode* on the *perceived depth* score. Three subsets were found as shown in Figure 7.4. The perceived amount of depth was lowest for the depth mode 'floating icon', while it was highest for the 'sinking background' depth mode.

Table 7.4 Results of a Tukey post-hoc test on the variable *depth mode* for the *perceived depth* scores for the full block analysis

Depth mode	Subset		
	1	2	3
Floating Icon	3.1464		
Window Effect		3.5589	
Sinking Background			3.8635
Sig.	1.000	1.000	1.000

The perceived amount of depth was higher for the natural background than for the uniform background. The mean score of the perceived amount of depth was μ = 3.61 with σ = 0.91 for the natural background and μ = 3.44 with σ = 0.93 for the uniform background. For the size of the 'selected icon', the perceived amount of depth was higher when the 'selected icon' was larger than the original icon size. The mean score of the perceived amount of depth for the bigger 'selected icon' was μ = 3.75 with σ = 0.86 and μ = 3.29 with σ = 0.93 for the original icon size.

A significant interaction on the *perceived depth* score was found for *background* and *shadow* ($F(1,423) = 6.133$, $p = 0.014$) and *shadow* and *size* ($F(1,423) = 4.546$, $p = 0.034$). An illustration of those interactions can be found in Figure 7.2. The effect of background and size on the perceived amount of depth is in both cases similar with or without shadow, but is more pronounced in the absence of a shadow.



(a) (b)

Figure 7.2 Illustration of the interaction on *perceived depth* between: (a) *background* and *shadow* and (b) *size* and *shadow*

7.5.2 Image quality

Incomplete design analysis

For the attribute *image quality*, an ANOVA was done with the *image quality score* as the dependent variable. Again, the *depth mode*, *background*, *shadow*, and *size* were used as fixed factors, including their two-way interaction and the subjects' *name* as random factor. The results are summarized in Appendix C, Table 11, and show that *depth mode* ($F(4,550) = 8.307$, $p < 0.001$) and *name* ($F(18,550) =$

10.623, $p < 0.001$) had a significant effect on the *image quality score*. The variables *background* ($F(2,550) = 2.340$, $p = 0.097$), *shadow* ($F(1,550) = 1.650$, $p = 0.199$), and *size* ($F(1,550) = 0.062$, $p = 0.803$) did not have a significant effect on the *image quality* score.

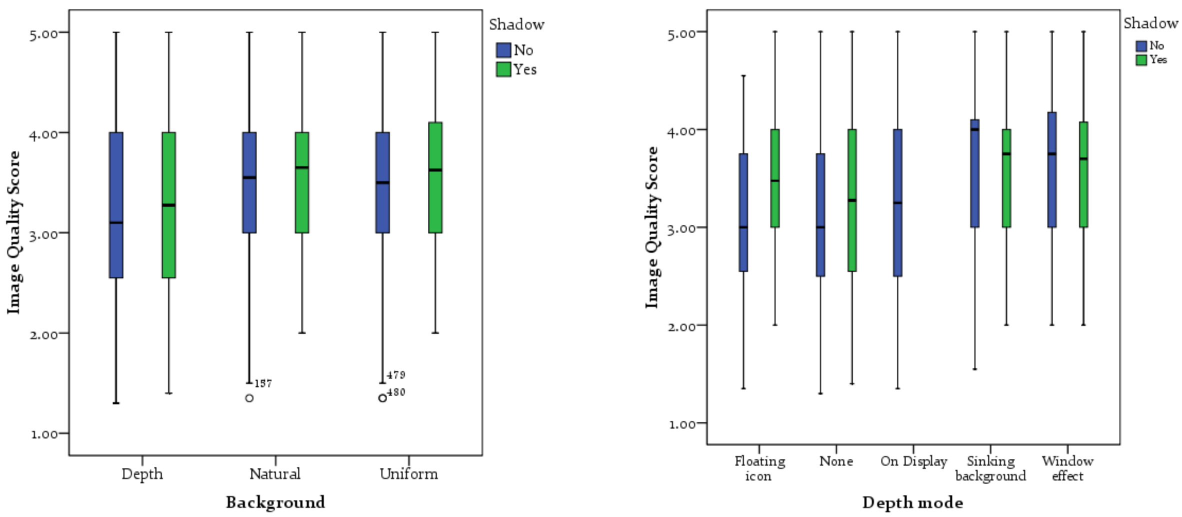
A Tukey post-hoc test on the variable *depth mode* revealed two subsets, as shown in Table 7.5. The image quality was significantly higher for the depth modes, in which the background was behind the screen than for the depth modes, in which the background contained depth or was at the screen. It should be noted that disparity as such did not necessarily have a (positive) effect on image quality, since the image quality score for the depth mode 'on display' was not significantly lower than for the depth modes 'floating icon' and 'none' (i.e. background with depth).

Table 7.5 Results of a Tukey post-hoc test on the variable *depth mode* for the scores on *image quality*

Depth mode	Subset	
	1	2
None	3.23	
On Display	3.24	
Floating Icon	3.29	
Window Effect		3.63
Sinking Background		3.64
Sig.	.966	1.000

From all two-way interactions between the fixed factors only the interaction between *background* and *shadow* ($F(1,550) = 3.987$, $p = 0.046$), and the interaction between *depth mode* and *shadow* ($F(2,550) = 3.168$, $p = 0.043$) had a significant effect on the image quality.

The interaction effect of *background* and *shadow* is depicted in Figure 7.3(a), here it is shown that the effect of the background on the image quality is the same for both shadow values. The image quality for the depth modes where the 'selected icon' has a disparity is higher when there is shadow added to the 'selected icon', while for the depth mode 'sinking background' the image quality is lower when shadow is added. For the depth mode 'window effect' the image quality is the same for both shadow values. This is illustrated in Figure 7.3(b).



(a) (b)

Figure 7.3 Illustration of the significant two-way interaction for the *image quality* scores: (a) the interaction between *background* and *shadow*, and (b) the interaction between *depth mode* and *shadow*

Full block design analysis

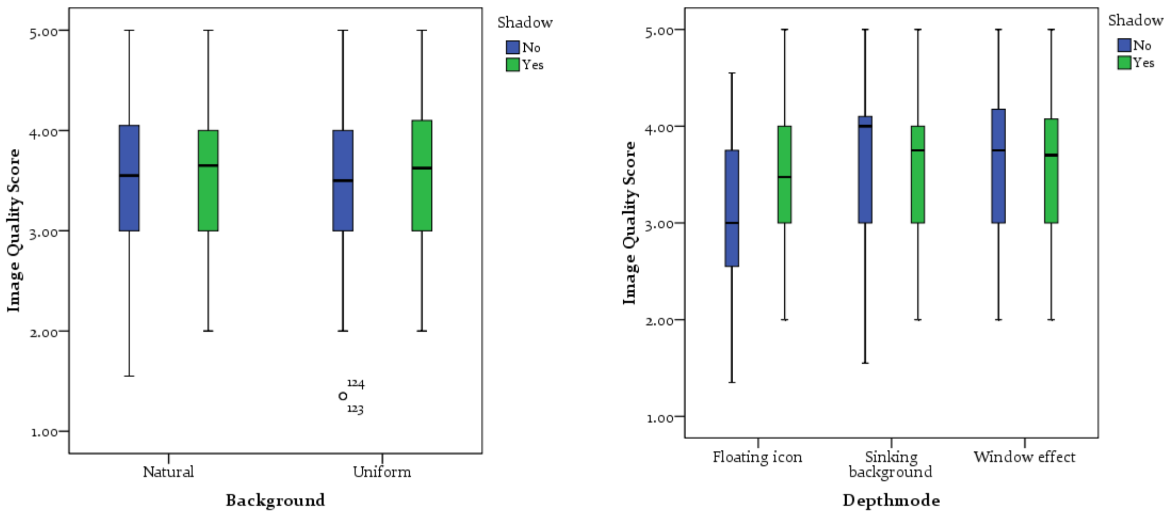
For the full block design analysis of the attribute *image quality*, an ANOVA was done with the *image quality score* as the dependent variable. Again, the *depth mode*, *background*, *shadow*, and *size* were used as fixed factors, including their two-way interaction and the subjects' *name* as random factor. The results are summarized in Appendix C, Table 12Table 11, and show that *depth mode* ($F(2,423) = 15.817$, $p < 0.001$) and *name* ($F(18,423) = 11.589$, $p < 0.001$) had a significant effect on the *image quality score*. The variables *background* ($F(1,423) = 2.019$, $p = 0.156$), *shadow* ($F(1,423) = 1.402$, $p = 0.237$), and *size* ($F(1,423) = 0.442$, $p = 0.507$) did not have a significant effect on the *image quality* score.

A Tukey post-hoc test on the variable *depth mode* revealed two subsets, as shown in Table 7.6. The image quality was significantly lower for the depth mode 'floating icon' than for the depth modes 'window effect' and 'sinking background'.

Table 7.6 Results of a Tukey post-hoc test on the variable *depth mode* for the scores on *image quality* for the full block analysis

Depth mode	Subset	
	1	2
Floating Icon	3.2924	
Window Effect		3.6280
Sinking Background		3.6414
Sig.	1.000	.980

Only the interaction between *background* and *shadow* ($F(1,423) = 4.943$, $p = 0.027$), and the interaction between *depth mode* and *shadow* ($F(2,423) = 3.928$, $p = 0.020$) had a significant effect on the image quality. In Figure 7.4 (a) the interaction between background and shadow is illustrated. Here it can be seen that the image quality is slightly better when there was shadow added to the 'selected icon' than for the 'selected icon' without shadow. In Figure 7.4 (b) the interaction between depth mode and shadow is depicted. The image quality for the depth mode 'floating icon' is higher when there is shadow added to the 'selected icon'. For the depth modes 'sinking background' and 'window effect' the image quality was higher in the absence of shadow.



(a) (b)

Figure 7.4 Illustration of the interaction on *image quality* between: (a) *background* and *shadow* and (b) *depth mode* and *shadow*

7.5.3 Affordance

Incomplete design analysis
Also for the attribute *affordance*, an ANOVA was performed with the *affordance score* as the dependent variable, the *depth mode*, *background*, *shadow*, and *size* as fixed factors, including their two-way interaction and the subjects' *name* as random factor. The results shown in Appendix C, Table 13 demonstrated a significant effect of *depth mode* ($F(4,550) = 56.434$, $p < 0.001$), *size* ($F(1,550) = 121.092$, $p < 0.001$) and *name* ($F(18,550) = 25.076$, $p < 0.001$) on the *affordance scores*. The variables *background* ($F(2,550) = 0.233$, $p = 0.792$) and *shadow* ($F(1,550) = 1.646$, $p = 0.200$) did not have a significant effect on the *affordance scores*.

To analyze the effect of the variable *depth mode* on the *affordance scores* a Tukey post-hoc test was performed. The results revealed three subsets, as can be seen in Table 7.7. The affordance for the depth mode 'on display' was lowest, while the affordance for the depth mode 'sinking background' was highest. For both depth modes the affordance score was significantly different from the scores for the depth modes 'floating icon', 'none', and 'window effect'.

Table 7.7 Results of a Tukey post-hoc test on the variable *depth mode* for the scores on *affordance*

Depthmode	Subset		
	1	2	3
On Display	1.72		
Floating Icon		3.34	
None		3.40	
Window Effect		3.54	
Sinking Background			3.83
Sig.	1.000	.201	1.000

The affordance was higher when the 'selected icon' was larger in size (with $\mu = 3.88$ and $\sigma = 0.85$) than when the 'selected icon' had the original size (with $\mu = 2.95$ and $\sigma = 1.07$).

From all the two-way interactions only the interaction between *background* and *shadow* ($F(1,550) = 7.301$, $p = 0.007$) was significant. Figure 7.5 shows that the affordance for the natural background was comparable, independent on whether or not a shadow was added to the 'selected icon'. The affordance for the background with depth and the uniform background was slightly higher when a shadow was added to the 'selected icon'.

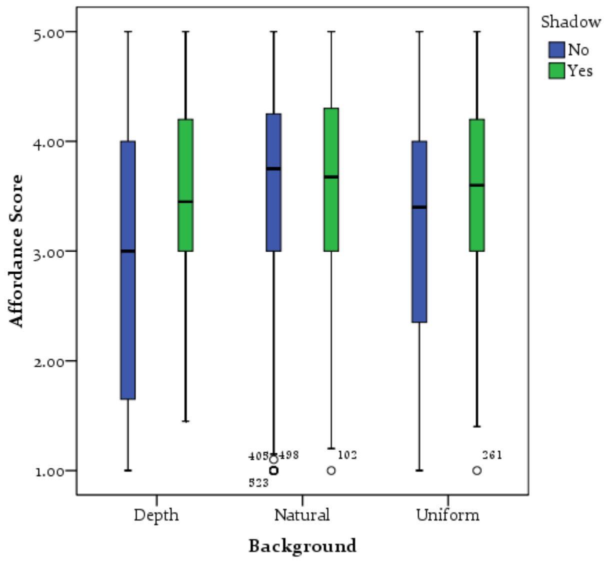


Figure 7.5 Illustration of the interaction between *background* and *shadow* on the *affordance* score

Full block design analysis

Also for the attribute *affordance*, an ANOVA was performed for the full block design analysis. With the *affordance score* as the dependent variable, the *depth mode*, *background*, *shadow*, and *size* as fixed factors, including their two-way interaction and the subjects' *name* as random factor. The results shown in Appendix C, Table 14, demonstrated a significant effect of *depth mode* ($F(2,423) = 24.394$, $p < 0.001$), *background* ($F(1,423) = 4.712$, $p = 0.031$), *size* ($F(1,423) = 128.659$, $p < 0.001$) and *name* ($F(18,423) = 24.182$, $p < 0.001$) on the *affordance scores*. The variable *shadow* ($F(1,423) = 0.102$, $p = 0.749$) did not have a significant effect on the *affordance scores*.

The effect of the variable *depth mode* on the affordance score was analyzed by performing a Tukey post-hoc test. In Table 7.8 the resulting three subsets are shown. The affordance for the depth mode 'floating icon' was significant lower than for the depth modes 'window effect' and 'sinking background'. The depth mode 'sinking background' had the highest affordance.

Table 7.8 Results of a Tukey post-hoc test on the variable *depth mode* for the scores on *affordance* for the full block analysis

Depth mode	Subset		
	1	2	3
Floating Icon	3.3365		
Window Effect		3.5359	
Sinking Background			3.8263
Sig.	1.000	1.000	1.000

The affordance was higher for the natural background (μ = 3.63 and σ = 0.96) than for the uniform background (μ = 3.50 and σ = 0.91). When the 'selected icon' was larger in size the affordance was higher (μ = 3.89 and σ = 0.86) than when the 'selected icon' had the original size (μ = 3.24 and σ = 0.89).

A significant interaction was only found between *background* and *shadow* ($F(1,423) = 8.245$, $p = 0.004$). The interaction is illustrated in Figure 7.6. The affordance was higher for the natural background when no shadow was added to the 'selected icon'. For the uniform background the affordance was slightly higher when shadow was added to the 'selected icon'.

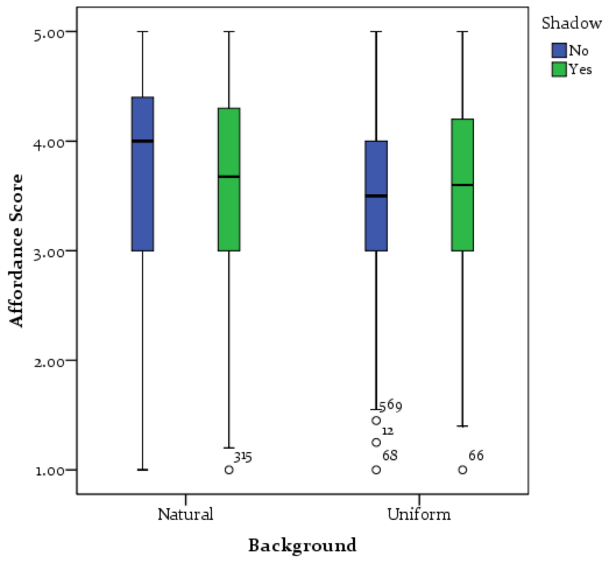


Figure 7.6 Illustration of the interaction between *background* and *shadow* on the *affordance* score

7.5.4 Aesthetics

Incomplete design analysis

Finally, an ANOVA was performed with the *aesthetic scores* as the dependent variable, again including the *depth mode*, *background*, *shadow*, and *size* as fixed factors, including their two-way interaction and the subjects' name as random factor. The results given in Appendix C, Table 15 illustrate that *depth mode* ($F(4,550) = 4.872$, $p = 0.001$), *background* ($F(2,550) = 5.334$, $p = 0.005$) and *name* ($F(18,550) = 8.110$, $p < 0.001$) had a significant effect on the *aesthetic scores*. The variables *shadow* ($F(1,550) = 1.492$, $p = 0.222$), and *size* ($F(1,550) = 0.047$, $p = 0.828$) did not have a significant effect on the *aesthetic scores*. For this attribute no significant interactions were found.

A Tukey post-hoc test on the variable *depth mode*, the results of which are given in Table 7.9, showed the existence of three subsets. The depth modes 'on display' and 'floating icon' had the lowest aesthetics scores. The aesthetic scores for the depth modes 'window effect' and 'sinking background' were higher, but were not statistically significantly different from the aesthetics scores for the depth mode 'floating icon'. The depth modes with the highest aesthetic score (though not statistically significantly different from the depth modes 'window effect' and 'sinking background') was 'none', referring to the background with the intrinsic depth.

Table 7.9 Results of the Tukey post-hoc test on the variable *depth mode* for the score of *aesthetics*

Depth mode	**Subset**		
	1	2	3
On Display	2.95		
Floating Icon	3.15	3.15	
Window Effect		3.28	3.28
Sinking Background		3.36	3.36
None			3.47
Sig.	.209	.173	.290

To analyze the effect of the variable *backgrounds* on the *scores of aesthetics*, a Tukey post-hoc test was performed. As can be seen in Table 7.10, two subsets were found. The aesthetics for the uniform background was significantly lower than for the natural background and the background with depth.

Table 7.10 Results of the Tukey post-hoc test on the variable *background* for the score of *aesthetics*

Background	Subset	
	1	2
Uniform	3.10	
Natural		3.35
Depth		3.42
Sig.	1.000	.630

Full block design analysis

For the full block design analysis, an ANOVA was performed with the *aesthetic scores* as the dependent variable, again including the *depth mode*, *background*, *shadow*, and *size* as fixed factors, including their two-way interaction and the subjects' name as random factor. The results given in Appendix C, Table 16, illustrate that *depth mode* ($F(2,423) = 4.777$, $p = 0.009$), *background* ($F(1,423) = 23.717$, $p < 0.001$) and *name* ($F(18,423) = 12.838$, $p < 0.001$) had a significant effect on the *aesthetic scores*. The variables *shadow* ($F(1,423) = 3.266$, $p = 0.071$), and *size* ($F(1,423) = 0.442$, $p = 0.507$) did not have a significant effect on the *aesthetic scores*. No significant interactions were found for this attribute.

A Tukey post-hoc test on the variable *depth mode* revealed two subsets, as shown in Table 7.11. The depth mode 'floating icon' had the lowest aesthetics score. 'Sinking background' was the depth mode with the highest aesthetic score (though not statistically significantly higher).

Table 7.11 Results of a Tukey post-hoc test on the variable *depth mode* for the scores on *aesthetics* for the full block analysis

Depth mode	Subset	
	1	2
Floating Icon	3.1474	
Window Effect	3.2839	3.2839
Sinking Background		3.3579
Sig.	.120	.533

The aesthetics score was higher for the natural background ($\mu = 3.40$ and $\sigma = 0.73$) than for the uniform background ($\mu = 3.13$ and $\sigma = 0.73$).

7.5.5 Relation between attributes

To evaluate whether the attributes used in the main experiment were mutually related or independent the Pearson correlation between all pairs of the four attributes was calculated. Table 7.12 shows that the highest correlation was found between the attributes *perceived depth* and *affordance* with a coefficient of $r = 0.683$, $N = 580$, $p < 0.001$. Although this correlation was statistically significant, its value was still moderate. In other words, there was a relation between the perceived amount of depth and the affordance, but the scores for the perceived amount of depth only explained about 40% of the variance in the affordance scores. Thus, affordance also contained other aspects. The correlation for the other combinations of attributes was considerably lower. More particularly, the correlation between *perceived amount of depth* and *image quality* was low. This confirmed some results reported in literature, showing that the presence or absence of stereoscopic depth had (almost) no impact on image quality (Seuntiëns, et al., 2005).

Table 7.12 Results of the Pearson correlation analysis on the four attributes: *perceived depth*, *image quality*, *affordance*, and *aesthetics*

Attributes			Pearson Correlation (r)	N	Significance (p)
Perceived depth	-	**Image quality**	0.262	589	0.000
Perceived depth	-	**Affordance**	0.683	589	0.000
Perceived depth	-	**Aesthetics**	0.245	589	0.000
Image quality	-	**Affordance**	0.207	589	0.000
Image quality	-	**Aesthetics**	0.309	589	0.000
Affordance	-	**Aesthetics**	0.298	589	0.000

7.6 Discussion

The results of the main experiment showed that the same trend exists for the incomplete experimental design and the full block design. This means that it was possible to analyze the data as though all variables were combined to each other.

From the results of the main experiment we can conclude that the depth mode was the most sensitive variable, since it was the only variable with a significant main effect on all four attributes. Adding shadow or not did not have a big effect on the overall experience: it did not have a significant effect on any of the attributes. It only showed some interaction with the choice of background on perceived amount of depth, image quality and affordance. The choice of the background had a significant effect on the perceived depth and on the aesthetic appearance. The size of the 'selected icon' significantly improved the perceived depth and affordance.

In general the depth mode 'sinking background' with the icon at the display screen and the background behind the display screen was most appreciated: it was scored high on perceived amount of depth, image quality, affordance and aesthetics. The depth mode 'floating icon' with the background at the display screen and the icon floating in front of the display screen was not highly appreciated: it was scored low on all four attributes. These findings are in line with earlier studies, in which it was found that the 'sinking background' depth mode had a better reading performance than the 'floating text' depth mode (Mizobuchi, et al., 2008).

The depth mode 'none' was most appreciated from the aesthetic point of view. Most subjects also confirmed during the experiment that they liked this background the most. Apparently, it was not the perceived amount of depth of this depth mode, nor its image quality that contributed to its aesthetic appearance, since both aspects were only scored low to average.

The depth mode 'on display' scored low on perceived amount of depth, which is logical since this depth mode corresponded to just a 2D representation of the user interface. This depth mode also scored significantly lower on image quality, affordance and aesthetics than most of the 3D depth modes, illustrating the added value of introducing stereoscopic depth in the user interface of a mobile device.

The choice of the background image did not significantly affect the assessment on image quality and aesthetics. It did, however, affect the assessment on perceived amount of depth and affordance. Especially the uniform background was not appreciated from an aesthetic point of view. The natural background had the highest score on the perceived amount of depth, and was also scored high from an aesthetic point of view. The linear perspective present in the natural image may have contributed to the perceived amount of depth.

As mentioned above, increasing the size of the ‘selected icon’ contributed to the perceived amount of depth and to the affordance. Apparently, the use of the relative size depth cue enhanced the depth perception as introduced by the stereoscopic depth. The increase in affordance implies that the bigger size of the icon made it easier to recognize that the icon had a function. The drawback of increasing the size of the ‘selected icon’ was that there was a small (but not statistically significant) reduction in image quality. In other words, when the size cue is used in the design of a user interface of a mobile device care should be taken to avoid any reduction in the quality rendering of the icon (for example by using vector images of icons).

The most significant interaction was the one between background and shadow: it had a significant effect on perceived amount of depth, the image quality and the affordance. In all cases the effect of the background on the attribute was strongest in the absence of a shadow added to the ‘selected icon’. Adding a shadow to the ‘selected icon’ reduced the difference between backgrounds for the various attributes.

Although this main experiment was a large and carefully designed experiment, it still had some limitations. The first limitation is in the group of participants. As discussed in the beginning of this chapter, a group of participants with a rather homogenous preference in disparity was selected. It might be that some of the conclusions drawn from this experiment deviate for people with a different preference in disparity. Additionally, it should be kept in mind that the experiment was based on an incomplete block design. In some cases the choices for all variables could not be combined from a practical point of view (for example in the case of the background with depth), and in some cases combinations of variables were omitted in order to keep the length of the experiment doable. It is not clear how the particular choice of the variables might have affected the scoring behavior of the participants.

8 Conclusions

The goal of this study was to find the added value of stereoscopic depth for the graphical user interface of a mobile device and how monocular depth cues combined with stereopsis influence the overall experience of the graphical user interface. The monocular depth cues used were shadow, relative size and relative brightness. Additional variables in the experiment were the background image of the graphical user interface and the depth mode referring to how background and icons were distributed in depth with respect to the display screen. The experience with a given user interface was measured using four attributes: perceived depth, image quality, affordance, and aesthetics. In total, three experiments were performed. In the first experiment, i.e. the pilot experiment, the attributes needed in the main experiment were established. The second experiment, i.e. a tuning experiment, aimed to find the preferred setting for the variables disparity, size and luminance. The last experiment, i.e. the main experiment, combined the results of the pilot and the tuning experiment to evaluate the design of a graphical user interface on a mobile device along the four attributes mentioned above. The most important results of all three experiments are summarized and discussed in this chapter, and recommendations for future works are given.

8.1 Added value of stereoscopic display

According to the results of this study a stereoscopic display has added value for the overall experience of a graphical user interface on a mobile device. In all four attributes the 'on display' depth mode, i.e. the 2D version of the graphical user interface, had the lowest mean scores. In other words, a stereoscopic rendering of the graphical user interface was preferred over a 2D rendering on perceived amount of depth, image quality, affordance and aesthetics. This conclusion, however, may have been biased by the specific selection of the subjects of the main experiment. These subjects were chosen for their similar and averaged preference in disparity for the stereoscopic rendering. Also including subjects with a preference for low disparity may reduce the added value of a stereoscopic display.

The amount of disparity that was used in the main experiment was about 1 cm in front of the display screen and 2 cm behind the display screen. There is a large spread in the preferred amount of disparity between the subjects. Therefore, it may be an option to allow the users to set their own preferred amount of disparity on their mobile device.

8.2 Attributes

The four attributes used in this study to evaluate the overall experience of a graphical user interface on a mobile device were perceived depth, image quality, affordance, and aesthetics. The results of this study showed that not all attributes scored high or low for the same combination of the variables depth mode, background image, size and shadow. This indicates that the attributes represent different assessment criteria. Only the attributes perceived depth and affordance seem to have some aspects in common, since the mutual correlation between these two attributes was relatively high, at least considerably higher than for any other pair of attributes.

8.3 Monocular depth cues

In addition to the stereoscopic cue also the monocular depth cues shadow, relative size and luminance were evaluated. The added value of each of these monocular depth cues is discussed in more detail below.

Shadow
The results of the pilot experiment have demonstrated that the added value of adding a shadow strongly depends on the shape of the shadow. In the pilot experiment, subjects commented that the shadow was too big and artificial, but that they would like a shadow when it was smaller and more realistic. Therefore, the shadow used in the main experiment was made smaller and more realistic. As a consequence, subjects scored the graphical user interfaces, in which the icons had a shadow, higher than the graphical user interfaces with icons without shadow on all four attributes. However, the difference in the scores for these four attributes was not statistically significant. Hence, there is a tendency that hypothesis 1, formulated as *"Users have an increased sense of depth when shadows are combined with stereopsis"* is true, but was confirmed in our study. Confirmation for the hypothesis might be found when either increasing the number of subjects, or by more carefully designing the size and shape of the shadow.

Relative size
Hypothesis 3 stating that *"Users have an increased sense of depth when changes in size are combined with stereopsis"* is true, since a change in size of the 'selected icon' yielded a higher score for the perceived amount of depth. The higher score for the perceived amount of depth was accompanied by a higher score for affordance, and so a change in size of the 'selected icon' also helps to understand the function of the 'selected icon'. Whether the change in size is also appreciated by the subjects from an overall experience point of view is less clear, since it did not contribute to higher scores for image quality and aesthetics.

Luminance
The results of the tuning experiment showed that the preferred luminance for most subjects was only 30% higher than the standard luminance used for the icons. As a consequence, adding luminance was not considered as very useful for the increased performance of a graphical user interface, since a luminance increase of 30% on a relatively high luminance is barely visible to the human eye, especially in the absence of a reference image. In earlier studies (Baldassi & Burr, 2004; Theeuwes & Lucassen, 1993), the authors showed some added value of luminance to perceived depth, but in these studies they used a set of equal objects and changed the luminance of only one of them. In other words, in their experimental setup, there was reference material available to compare the luminance of one object with. In our experiment the six icons on the graphical user interface were not the same, and so, one icon could not be directly compared to another one as reference for the luminance. Hence, hypothesis 2 stating that *"Users have an increased sense of depth when luminance differences are combined with stereopsis"* cannot be confirmed, nor rejected, since it was not tested in the main experiment. Nonetheless, one can wonder whether adding luminance to one icon would have any effect in a graphical user interface that has so many different backgrounds and icons.

Combination of cues
Hypothesis 4 states that *"Users have an even larger sense of depth when shadows, luminance differences and changes in size are combined all together with stereopsis"*. This hypothesis can only be partly answered, namely only for the combination of the monocular cues shadow and relative size, since the variable shadow has no significant effect on the attributes, the sense of depth when the variables shadow and size are combined is solely dependent on the variable size. Therefore, hypothesis 4 is not true.

8.4 Background

The use of different backgrounds only had a significant effect on the scores for perceived depth and aesthetics. The use of a natural background scored highest on perceived depth, and also on affordance, but for the latter the difference with the other backgrounds was not statistically significant. The uniform background scored highest on image quality, but again the difference in score with the other backgrounds was not statistically significant. The background with depth scored significantly highest on aesthetics. Hence, hypothesis 5, namely that *"Users have a stronger sense of depth when using a natural instead of a uniform background",* is true, since the natural background had the highest score for the perceived amount of depth. The confirmation of this hypothesis is in contradiction to the results of Arikan et al. where the different backgrounds did not have a significant effect on the perceived amount of depth (Arikan, et al., 2008). However, in our study there was more variation in backgrounds than in the study of Arikan et al.. Whether the natural background also contributes to the overall appreciation of the graphical user interface is less clear, since it does not result in higher scores for image quality or aesthetics. The background with depth is most appreciated from an aesthetic point of view, but apparently it does not significantly contribute to the perceived amount of depth.

8.5 Depth mode

The depth mode had a significant effect on all four attributes. The depth mode 'sinking background', in which the icon is at the screen and the background behind the screen, had the highest score on the perceived amount of depth, image quality and affordance. Hence, hypothesis 6 stating that *"Users report a preferred experience when the selected icon of the graphical user interface is 'on the screen', while the other icons are placed 'behind the screen' "*, is true. On aesthetics the depth mode 'none', in which the background already had some depth and the icon was on the screen, had the highest score, followed by the depth mode 'sinking background'. So, also from the point of view of aesthetics hypothesis 6 is confirmed.

In agreement with the paper by (Mizobuchi, et al., 2008) the depth mode 'floating icon', in which the background was at the screen and the icon floating in front of the screen, was not preferred at all. It scored low on perceived amount of depth, image quality, affordance and aesthetics. The fact that the icon becomes blurred when it is rendered in front of the screen might explain the low score on image quality and aesthetics. So, this mode is not at all recommended for the design of a graphical user interface on a mobile device.

8.6 Overall experience

The overall experience of the graphical user interface was defined as the attributes perceived amount of depth, image quality, affordance and aesthetics. When the perceived amount of depth was high the affordance was high as well. The attributes image quality and aesthetics, however, were not high when the perceived amount of depth was high. Hypothesis 7 states that *"Users better appreciate the overall experience of a graphical user interface when they experience an increased sense of depth"*. This hypothesis is only true for affordance, as image quality and aesthetics did not increase when the sense of depth increased.

8.7 Recommendations and future works

The findings in this study indicate that monocular depth cues and the choice of depth mode and background image can be used in addition to stereopsis to enhance the overall experience of a graphical user interface of a mobile device. However, not all hypotheses for this study were conclusive. More experiments are needed for the variables shadows, luminance and size, especially to determine their optimal setting more accurately. This can be done via an experiment in which stereopsis is combined with various shadow sizes and shades in combination with more different backgrounds. In tour current experiment only one natural background, one uniform background and one background with depth was used. More variation in the background image would allow evaluation of the effect of various colors in the natural background on the preferred shade for the shadow of the 'selected icon'. The size of the shadow is an important factor as well, as was voiced by the subjects in the pilot experiment; it should not be too obvious. Including more variation in size and shade of the shadow should allow us to draw a more conclusive conclusion about the sense of depth experienced when shadows are combined with stereopsis. Additionally, it should be investigated more accurately which luminance differences can be actually seen in a graphical user interface, especially in view of the fact that graphical user interfaces usually are very colorful, because of the background and the different icons. Based on the results it can then be decided whether it makes sense to add a luminance difference to the 'selected icon' in order to make it more popping out from the background. The added value of increasing the size of the 'selected icon' can be studied further by using icons of better quality. As mentioned earlier, the icons used in this study were only available in png format. This meant that when the icons were blown up, the quality of the icons deteriorated.

In the current study only one natural background and one background with depth was used. The content of both background images was different. As a consequence, we could not conclude whether the higher aesthetic scores for the background with depth was due to the image content itself or to the fact that there was depth in the background. Therefore, an additional experiment is needed in which background images having the same content, but once rendered without depth and once rendered with depth are scored on aesthetics. This experiment should allow us to give an answer to whether the image content or the depth in the background affects the aesthetics of the graphical user interface most.

One of the most dominant monocular depth cues is known to be occlusion (Cutting & Vishton, 1995). This cue was not explicitly used in the current design of the graphical user interface of the mobile device. It might be interesting to see whether occlusion can have added value to the sense of depth and overall experience of a graphical user interface.

A simplified user interface was used in this study. It existed of a background with only six icons on top of it, of which only one icon was displayed as being selected. It would be to repeat the current study with a more realistic user interface of a mobile device; such that subjects get the feeling that they are looking at a real user interface. Additionally, it would be more realistic, and probably make a big difference to the results when a working prototype of a graphical user interface based on an auto-stereoscopic display could be used. In our current study no interaction with the user interface was possible. In case interaction would be possible, the evaluation of e.g. affordance of the graphical user interface, and consequently, also of its overall experience would be more reliable. With a real interactive system also more usability tests could be performed.

This study was limited to using static images only. However, nowadays most user interfaces of mobile devices use animation. It might be interesting to see if animation has an additional value to the overall experience of a graphical user interface, and more particularly whether the effect of introducing stereopsis and monocular depth cues is stronger with animations in the graphical user interface than when using static images only. Animations on an auto-stereoscopic display are not limited by horizontal or vertical motions only, but can also make use of the depth dimension. Based on the results of this study, it is expected that animations behind the display screen are more appreciated than animations that come too far out to the front of the screen. Besides the visual part of the depth experience, new interactions with 3D displays might enhance the depth experience. Interactions for 3D displays that are more natural and intuitive can make the mobile device easier to use, as can be expected from literature on computers (Liu, Pastoor, Seifert, & Hurtienne, 2000).

References

Arikan, S., Bos, L., & Wu, S. L. (2008). Depth perception with auto-stereoscopic displays. Unpublished Image Quality Course Report. Delft University of Technology.

Baldassi, S., & Burr, D. C. (2004). "Pop-out" of targets modulated in luminance or colour: the effect of intrinsic and extrinsic uncertainty. *Vision Research, 44*(12), 1227-1233.

Bay, S., & Ziefle, M. (2005). Children Using Cellular Phones: The Effects of Shortcomings in User Interface Design. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 47*, 158-168.

Boydstun, A. S., Rogers, J. A., Tripp, L. M., & Patterson, R. (2009). Stereoscopic depth perception survives significant interocular luminance differences. *Journal of the Society for Information Display, 17*(5), 467-471.

Cockburn, A., & McKenzie, B. (2001). *3D or not 3D?: evaluating the effect of the third dimension in a document management system*. Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.

Cockburn, A., & McKenzie, B. (2002). *Evaluating the effectiveness of spatial memory in 2D and 3D physical and virtual environments*. Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves.

Cockburn, A., & McKenzie, B. (2004). Evaluating spatial memory in two and three dimensions. *International Journal of Human-Computer Studies, 61*(3), 359-373.

Cutting, J., & Vishton, P. (1995). Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In W. Epstein & S. Rogers (Eds.), *Perception of Space and Motion (Handbook of Perception and Cognition)* (2nd ed., pp. 69-117). New York: Academic Press.

Dix, A., Finlay, J. E., Abowd, G. D., & Beale, R. (2003). *Human-Computer Interaction* (3rd ed.): Prentice Hall.

Dodgson, N. A. (2005). Autostereoscopic 3D Displays. *Computer, 38*(8), 31-36.

Dunlop, M., & Brewster, S. (2002). The Challange of Mobile Devices for Human Computer Interaction. *Personal and Ubiquitous Computing, 6*(4), 235-236.

Felisberti, F. M., Solomon, J. A., & Morgan, M. J. (2005). The role of target salience in crowding. *Perception, 34*(7), 823-833.

Geib, T., & Baumann, C. (1990). Effect of luminance and contrast on stereoscopic acuity. *Graefe's Archive for Clinical and Experimental Ophthalmology, 228*(4), 310-315.

Gong, J., & Tarasewich, P. (2004). *Guidelines for Handheld Mobile Device Interface Design.* Paper presented at the Proceedings of the Decision Sciences Institute (DSI) Annual Meeting, Boston, Massachusetts

Hakala, T., Lehikoinen, J., & Aaltonen, A. (2005). *Spatial interactive visualization on small screen*. Paper presented at the Proceedings of the 7th international conference on Human computer interaction with mobile devices & services.

Häkkinen, J., Takatalo, J., Kilpeläinen, M., Salmimaa, M., & Nyman, G. (2009). Determining limits to avoid double vision in an autostereoscopic display: Disparity and image element width. *Journal of the Society for Information Display, 17*(5), 433-441.

Heynderickx, I. E. J. R. (2007). *3D display technology: an overview.* Paper presented at the LatinDisplay 2007, Brazil.

Hoffman, D. M., Girshick, A. R., Akeley, K., & Banks, M. S. (2008). Vergence-accommodation conflicts hinder visual performance and cause visual fatigue. *Journal of Vision, 8*(3), 1-30.

IJsselsteijn, W., de Ridder, H., Hamberg, R., Bouwhuis, D., & Freeman, J. (1998). Perceived depth and the feeling of presence in 3DTV. *Displays, 18*, 207-214.

ITU. (2009). International Telecommunication Union. Retrieved 7 September, 2009, from http://www.itu.int/ITU-D/ICTEYE/Indicators/Indicators.aspx

Kaikkonen, A., Kekäläinen, A., Cankar, M., Kallio, T., & Kankainen, A. (2005). Usability Testing of Mobile Applications: A Comparison between Laboratory and Field Testing. *Journal of Usability Studies, 1*(1), 4-16.

Karlsson, P., & Djabri, F. (2001). *Analogue styled user interfaces: An exemplified set of principles intended to improve aesthetic qualities in use.* Paper presented at the Proceedings of Mobile HCI 2001: Third International Workshop on Human-Computer Interaction with Mobile Devices.

Kiljander, H. (2004). *Evolution and Usability of Mobile Phone Interaction Styles.* Helsinki University of Technology, Espoo.

Klockar, T., Carr, D. A., Hedman, A., Johansson, T., & Bengtsson, F. (2003). *Usability of mobile phones.* Paper presented at the Proceeding of the 19th International Symposium on Human Factors in Telecommunications.

Kooi, F. L., & Toet, A. (2004). Visual comfort of binocular and 3D displays. *Displays, 25*(2-3), 99-108.

Lambooij, M., IJsselsteijn, W., Fortuin, M., & Heynderickx, I. (2009). Visual Discomfort and Visual Fatigue of Stereoscopic Displays: A Review. *Journal of Imaging Science and Technology, 53*(3), 1-14.

Lambooij, M. T. M., IJsselsteijn, W. A., & Heynderickx, I. (2007). *Visual discomfort in stereoscopic displays: a review*.

Liu, J., Pastoor, S., Seifert, K., & Hurtienne, J. (2000). *Three-dimensional PC: toward novel forms of human-computer interaction.* Paper presented at the Three-Dimensional Video and Display: Devices and Systems SPIE, Boston.

Mather, G., & Smith, D. R. R. (2002). Blur discrimination and its relation to blur-mediated depth perception. *Perception, 31*(10), 1211-1219.

Mizobuchi, S., Terasaki, S., Hakkinen, J., Heinonen, E., Bergquist, J., & Chignell, M. (2008). The effect of stereoscopic viewing in a word-search task with a layered background. *Journal of the Society for Information Display, 16*(11), 1105-1113.

Nielsen, J. (1994). *Enhancing the explanatory power of usability heuristics*. Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems: celebrating interdependence.

NTTDocomo. (2002). Latest i-shot Model Offers World's First 3D Mobile Phone Screen. Retrieved 11 November, 2009, from http://www.nttdocomo.com/pr/2002/000907.html

Ostrovsky, Y., Cavanagh, P., & Sinha, P. (2005). Perceiving illumination inconsistencies in scenes. *Perception, 34*(11), 1301-1314.

Patterson, R., & Silzars, A. (2009). Immersive stereo displays, intuitive reasoning, and cognitive engineering. *Journal of the Society for Information Display, 17*(5), 443-448.

Rajae-Joordens, R. J. E. (2008). Measuring Experiences in Gaming and TV Applications. In J. H. D. M. Westerink, M. Ouwerkerk, T. J. M. Overbeek, W. F. Pasveer & B. de Ruyter (Eds.), *Probing Experience* (pp. 77-90): Springer Netherlands.

Redmond-Pyle, D., & Moore, A. (1995). *Graphical User Interface Design and Evaluation (GUIDE)* (1st ed.): Prentice-Hall.

Robertson, G., Czerwinski, M., Larson, K., Robbins, D. C., Thiel, D., & Dantzich, M. v. (1998). *Data mountain: using spatial memory for document management*. Paper presented at the Proceedings of the 11th annual ACM symposium on User interface software and technology.

Robertson, G., van Dantzich, M., Robbins, D., Czerwinski, M., Hinckley, K., Risden, K., et al. (2000). *The Task Gallery: a 3D window manager.* Paper presented at the CHI '00: Proceedings of the SIGCHI conference on Human factors in computing systems.

Salmimaa, M., Häkkinen, J., Liinasuo, M., & Järvenpää, T. (2009). Effect of number of views to the viewing experience with autostereoscopic 3-D displays. *Journal of the Society for Information Display, 17*(5), 449-458.

Seuntiëns, P. J., Heynderickx, I. E., IJsselsteijn, W. A., van den Avoort, P. M. J., Berentsen, J., Dalm, I. J., et al. (2005). *Viewing experience and naturalness of 3D images.* Paper presented at the Three-Dimensional TV, Video, and Display IV Boston, MA, USA.

Sharp. (2003). NTT DoCoMo launches second Sharp 3D phone in Japan. Retrieved 11 November, 2009, from http://www.sle.sharp.co.uk/news/newsitem.php?newsid=109

Shneiderman, B., & Plaisant, C. (2009). *Designing the User Interface: Strategies for Effective Human-Computer Interaction* (5th ed.): Addison Wesley.

Sparre, E. (2007). What to Look For in User Interface Toolkits for Mobile Phones. *TAT White Paper series*(TAT-ES-NR001-Q107).

Theeuwes, J., & Lucassen, M. P. (1993). An adaptation-induced pop-out in visual search. *Vision Research, 33*(16), 2353-2357.

Ware, C. (2004). *Information Visualization: Perception for design* (2nd ed.). San Francisco: Morgan Kaufmann.

Ware, C., & Franck, G. (1996). Evaluating stereo and motion cues for visualizing information nets in three dimensions. *ACM Transactions on Graphics, 15*(2), 121-140.

Wheatstone, C. (1838). Contributions to the Physiology of Vision. Part the First. On Some Remarkable, and Hitherto Unobserved, Phenomena of Binocular Vision. *Philosophical Transactions of the Royal Society of London, 128*, 371-394.

Yeh, Y. Y., & Silverstein, L. D. (1990). Limits of fusion and depth judgment in stereoscopic color displays. *Human Factors, 32*(1), 45-60.

Appendix A Tables of the pilot experiment

Table 1 List of all used combinations of variables

Image #	Stereopsis	Sinking background (SB)/Floating Icon(FI)/On display (OD)	Uniform/ Natural	Shadow	Luminance	Big Size (BS)/Small Size (SS)
Image X	-	OD	Uniform	-	-	SS
Image Y	-	OD	Natural	-	-	SS
Image 1	X	SB	Uniform	X	X	BS
Image 2	X	SB	Uniform	X	X	SS
Image 3	X	SB	Uniform	X	-	BS
Image 4	X	SB	Uniform	X	-	SS
Image 5	X	SB	Uniform	-	X	BS
Image 6	X	SB	Uniform	-	X	SS
Image 7	X	SB	Uniform	-	-	BS
Image 8	X	SB	Uniform	-	-	SS
Image 9	X	SB	Natural	X	X	BS
Image 10	X	SB	Natural	X	X	SS
Image 11	X	SB	Natural	X	-	BS
Image 12	X	SB	Natural	X	-	SS
Image 13	X	SB	Natural	-	X	BS
Image 14	X	SB	Natural	-	X	SS
Image 15	X	SB	Natural	-	-	BS
Image 16	X	SB	Natural	-	-	SS
Image 17	X	FT	Uniform	X	X	BS
Image 18	X	FT	Uniform	X	X	SS
Image 19	X	FT	Uniform	X	-	BS
Image 20	X	FT	Uniform	X	-	SS
Image 21	X	FT	Uniform	-	X	BS
Image 22	X	FT	Uniform	-	X	SS
Image 23	X	FT	Uniform	-	-	BS
Image 24	X	FT	Uniform	-	-	SS
Image 25	X	FT	Natural	X	X	BS
Image 26	X	FT	Natural	X	X	SS
Image 27	X	FT	Natural	X	-	BS
Image 28	X	FT	Natural	X	-	SS
Image 29	X	FT	Natural	-	X	BS
Image 30	X	FT	Natural	-	X	SS
Image 31	X	FT	Natural	-	-	BS
Image 32	X	FT	Natural	-	-	SS

Table 2 ANOVA of the main effects and two-way interactions of the dependent variable *affordance*, with *depth mode*, *background*, *shadow*, *luminance* and *size* as fixed factors and *name* as random factor

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Depth mode	36.351	2	18.175	46.298	.000
Background	.081	1	.081	.207	.649
Shadow	1.563	1	1.563	3.982	.047
Luminance	.013	1	.013	.032	.858
Size	47.672	1	47.672	121.437	.000
Name	148.203	11	13.472	34.320	.000
Depth mode x Background	1.458	2	.729	1.857	.157
Background x Luminance	.002	1	.002	.004	.948
Background x Shadow	3.064	1	3.064	7.804	.005
Background x Size	.016	1	.016	.041	.839
Depth mode x Luminance	.100	1	.100	.255	.614
Depth mode x Shadow	1.138	1	1.138	2.898	.090
Depth mode x Size	.022	1	.022	.056	.813
Shadow x Luminance	.113	1	.113	.289	.591
Luminance x Size	.338	1	.338	.862	.354
Shadow x Size	.016	1	.016	.041	.839
Error	148.783	379	.393		

Table 3 ANOVA of the main effects and two-way interactions of the dependent variable *aesthetics*, with *depth mode*, *background*, *shadow*, *luminance* and *size* as fixed factors and *name* as random factor

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Depth mode	34.621	2	17.311	43.088	.000
Background	.708	1	.708	1.761	.185
Shadow	4.115	1	4.115	10.242	.001
Luminance	.195	1	.195	.485	.487
Size	6.342	1	6.342	15.787	.000
Name	92.510	11	8.410	20.934	.000
Depth mode x Background	.636	2	.318	.792	.454
Background x Luminance	.086	1	.086	.214	.644
Background x Shadow	2.446	1	2.446	6.089	.014
Background x Size	.029	1	.029	.073	.788
Depth mode x Luminance	.347	1	.347	.865	.353
Depth mode x Shadow	.145	1	.145	.360	.549
Depth mode x Size	.023	1	.023	.056	.812
Shadow x Luminance	.089	1	.089	.222	.638
Luminance x Size	.126	1	.126	.313	.576
Shadow x Size	.004	1	.004	.010	.920
Error	152.262	379	.402		

Table 4 ANOVA of the main effects and two-way interactions of the dependent variable *preference*, with *depth mode*, *background*, *shadow*, *luminance* and *size* as fixed factors and *name* as random factor

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Depth mode	70.284	2	35.142	70.369	.000
Background	1.055	1	1.055	2.113	.147
Shadow	.160	1	.160	.321	.571
Luminance	.430	1	.430	.861	.354
Size	31.425	1	31.425	62.925	.000
Name	35.111	11	3.192	6.392	.000
Depth mode x Background	.475	2	.238	.476	.622
Background x Luminance	.061	1	.061	.123	.726
Background x Shadow	.043	1	.043	.086	.770
Background x Size	.122	1	.122	.245	.621
Depth mode x Luminance	.148	1	.148	.297	.586
Depth mode x Shadow	.005	1	.005	.010	.922
Depth mode x Size	1.198	1	1.198	2.399	.122
Shadow x Luminance	.307	1	.307	.614	.434
Luminance x Size	.052	1	.052	.103	.748
Shadow x Size	1.036	1	1.036	2.075	.151
Error	189.272	379	.499		

Appendix B Tables of the tuning experiment

Table 5 ANOVA of the main effects and two-way interactions of the dependent variable *disparity*, with *depth mode* and *background* as fixed factors and *name* as random factor

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Depth mode	28842.081	2	14421.041	50.443	.000
Background	4299.680	1	4299.680	15.040	.000
Name	37930.672	36	1053.630	3.685	.000
Depth mode * Background	690.712	2	345.356	1.208	.301
Error	61752.139	216	285.890		

Table 6 ANOVA of the main effects and two-way interactions of the dependent variable *size*, with *disparity* and *depth mode* as fixed factors and *name* as random factor

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Disparity	691.892	1	691.892	2.246	.137
Depth mode	67.568	1	67.568	.219	.640
Name	39856.757	36	1107.132	3.594	.000
Disparity * Depth mode	172.973	1	172.973	.562	.455
Error	33267.568	108	308.033[b]		

Table 7 ANOVA of the main effects and two-way interactions of the dependent variable *luminance*, with *background* and *disparity* as fixed factors and *name* as random factor

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Background	4.054	1	4.054	.019	.892
Disparity	.901	2	.450	.002	.998
Name	45920.721	36	1275.576	5.828	.000
Background * Disparity	213.514	2	106.757	.488	.615
Error	39398.198	180	218.879		

Appendix C Tables of the main experiment

Table 8 Table of the result of the Tukey Post Hoc test for the variable *name*, with the preferred *disparity* as dependent variable. *Name*, *depth mode*, and *background* were the fixed factors. The green area indicates the subjects who participated in the main experiment

Name	Subset							
	1	2	3	4	5	6	7	8
Subject 1	36.00							
Subject 2	40.29	40.29						
Subject 3	40.29	40.29						
Subject 4	43.14	43.14	43.14					
Subject 5	45.29	45.29	45.29	45.29				
Subject 6	49.57	49.57	49.57	49.57	49.57			
Subject 7	49.57	49.57	49.57	49.57	49.57			
Subject 8	50.29	50.29	50.29	50.29	50.29	50.29		
Subject 9	50.29	50.29	50.29	50.29	50.29	50.29		
Subject 10	51.71	51.71	51.71	51.71	51.71	51.71		
Subject 11	53.86	53.86	53.86	53.86	53.86	53.86	53.86	
Subject 12	53.86	53.86	53.86	53.86	53.86	53.86	53.86	
Subject 13	54.57	54.57	54.57	54.57	54.57	54.57	54.57	54.57
Subject 14	54.57	54.57	54.57	54.57	54.57	54.57	54.57	54.57
Subject 15	55.29	55.29	55.29	55.29	55.29	55.29	55.29	55.29
Subject 16	56.00	56.00	56.00	56.00	56.00	56.00	56.00	56.00
Subject 17	59.57	59.57	59.57	59.57	59.57	59.57	59.57	59.57
Subject 18	60.29	60.29	60.29	60.29	60.29	60.29	60.29	60.29
Subject 19	61.71	61.71	61.71	61.71	61.71	61.71	61.71	61.71
Subject 20	63.14	63.14	63.14	63.14	63.14	63.14	63.14	63.14
Subject 21		65.00	65.00	65.00	65.00	65.00	65.00	65.00
Subject 22		66.00	66.00	66.00	66.00	66.00	66.00	66.00
Subject 23			69.29	69.29	69.29	69.29	69.29	69.29
Subject 24			69.57	69.57	69.57	69.57	69.57	69.57
Subject 25			69.57	69.57	69.57	69.57	69.57	69.57
Subject 26			70.29	70.29	70.29	70.29	70.29	70.29
Subject 27			70.71	70.71	70.71	70.71	70.71	70.71
Subject 28				71.43	71.43	71.43	71.43	71.43
Subject 29				71.71	71.71	71.71	71.71	71.71
Subject 30				71.71	71.71	71.71	71.71	71.71
Subject 31				73.14	73.14	73.14	73.14	73.14
Subject 32				73.14	73.14	73.14	73.14	73.14
Subject 33					73.86	73.86	73.86	73.86
Subject 34					75.29	75.29	75.29	75.29
Subject 35						77.86	77.86	77.86
Subject 36							80.29	80.29
Subject 37								82.14
Sig.	.077	.131	.065	.058	.131	.065	.101	.065

Table 9 ANOVA of the main effects and two-way interactions of the dependent variable *perceived depth scores*, with *depth mode*, *background*, *shadow*, and *size* as fixed factors and *name* as random factor

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Depth mode	157.770	4	39.442	96.947	.000
Background	2.720	2	1.360	3.342	.036
Shadow	.986	1	.986	2.424	.120
Size	24.475	1	24.475	60.157	.000
Name	160.007	18	8.889	21.849	.000
Depth mode x Background	.933	3	.311	.765	.514
Background x Shadow	2.197	1	2.197	5.399	.021
Background x Size	.313	1	.313	.770	.381
Depth mode x Shadow	.705	2	.352	.866	.421
Depth mode x Size	1.095	2	.548	1.346	.261
Shadow x Size	1.025	1	1.025	2.519	.113
Error	223.766	550	.407		

Table 10 ANOVA of the main effects and two-way interactions of the dependent variable *perceived depth scores*, with *depth mode*, *background*, *shadow*, and *size* as fixed factors and *name* as random factor. Only the depth modes: 'sinking background', 'floating icon', and 'window effect' are taken into account.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Depth mode	39.377	2	19.689	54.964	.000
Background	3.518	1	3.518	9.820	.002
Shadow	.293	1	.293	.817	.367
Size	23.925	1	23.925	66.790	.000
Name	161.280	18	8.960	25.013	.000
Depth mode x Background	.634	2	.317	.885	.413
Background x Shadow	2.197	1	2.197	6.133	.014
Background x Size	.313	1	.313	.874	.350
Depth mode x Shadow	.705	2	.352	.984	.375
Depth mode x Size	1.095	2	.548	1.528	.218
Shadow x Size	1.628	1	1.628	4.546	.034
Error	151.523	423	.358		

Table 11 ANOVA of the main effects and two-way interactions of the dependent variable *image quality scores*, with *depth mode*, *background*, *shadow*, and *size* as fixed factors and *name* as random factor

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Depth mode	15.477	4	3.869	8.307	.000
Background	2.179	2	1.090	2.340	.097
Shadow	.769	1	.769	1.650	.199
Size	.029	1	.029	.062	.803
Name	89.058	18	4.948	10.623	.000
Depth mode x Background	.827	3	.276	.592	.621
Background x Shadow	1.857	1	1.857	3.987	.046
Background x Size	.626	1	.626	1.345	.247
Depth mode x Shadow	2.951	2	1.476	3.168	.043
Depth mode x Size	.446	2	.223	.479	.619
Shadow x Size	.106	1	.106	.227	.634
Error	256.165	550	.466		

Table 12 ANOVA of the main effects and two-way interactions of the dependent variable *image quality scores*, with *depth mode*, *background*, *shadow*, and size as fixed factors and *name* as random factor. Only the depth modes: 'sinking background', 'floating icon', and 'window effect' are taken into account.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Depth mode	11.885	2	5.942	15.817	.000
Background	.759	1	.759	2.019	.156
Shadow	.527	1	.527	1.402	.237
Size	.166	1	.166	.442	.507
Name	78.375	18	4.354	11.589	.000
Depth mode x Background	.194	2	.097	.258	.773
Background x Shadow	1.857	1	1.857	4.943	.027
Background x Size	.626	1	.626	1.667	.197
Depth mode x Shadow	2.951	2	1.476	3.928	.020
Depth mode x Size	.446	2	.223	.594	.552
Shadow x Size	.127	1	.127	.337	.562
Error	158.923	423	.376		

Table 13 ANOVA of the main effects and two-way interactions of the dependent variable *affordance scores*, with *depth mode*, *background*, *shadow*, and *size* as fixed factors and *name* as random factor

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Depth mode	96.362	4	24.091	56.434	.000
Background	.199	2	.100	.233	.792
Shadow	.703	1	.703	1.646	.200
Size	51.692	1	51.692	121.092	.000
Name	192.679	18	10.704	25.076	.000
Depth mode x Background	1.049	3	.350	.819	.484
Background x Shadow	3.117	1	3.117	7.301	.007
Background x Size	.042	1	.042	.099	.753
Depth mode x Shadow	.157	2	.078	.183	.833
Depth mode x Size	.279	2	.140	.327	.721
Shadow x Size	.157	1	.157	.369	.544
Error	234.786	550	.427		

Table 14 ANOVA of the main effects and two-way interactions of the dependent variable *affordance scores*, with *depth mode*, *background*, *shadow*, and *size* as fixed factors and *name* as random factor. Only the depth modes: 'sinking background', 'floating icon', and 'window effect' are taken into account.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Depth mode	18.443	2	9.222	24.394	.000
Background	1.781	1	1.781	4.712	.031
Shadow	.039	1	.039	.102	.749
Size	49.014	1	49.014	129.659	.000
Name	164.546	18	9.141	24.182	.000
Depth mode x Background	.204	2	.102	.270	.763
Background x Shadow	3.117	1	3.117	8.245	.004
Background x Size	.042	1	.042	.112	.738
Depth mode x Shadow	.157	2	.078	.207	.813
Depth mode x Size	.279	2	.140	.369	.692
Shadow x Size	.190	1	.190	.502	.479
Error	159.903	423	.378		

Table 15 ANOVA of the main effects and two-way interactions of the dependent variable *aesthetics scores*, with *depth mode*, *background*, *shadow*, and *size* as fixed factors and *name* as random factor

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Depth mode	8.680	4	2.170	4.872	.001
Background	4.752	2	2.376	5.334	.005
Shadow	.665	1	.665	1.492	.222
Size	.021	1	.021	.047	.828
Name	65.017	18	3.612	8.110	.000
Depth mode x Background	1.565	3	.522	1.171	.320
Background x Shadow	.308	1	.308	.691	.406
Background x Size	.058	1	.058	.131	.718
Depth mode x Shadow	.585	2	.292	.656	.519
Depth mode x Size	.298	2	.149	.335	.715
Shadow x Size	.055	1	.055	.123	.726
Error	244.966	550	.445		

Table 16 ANOVA of the main effects and two-way interactions of the dependent variable *aesthetics scores*, with *depth mode*, *background*, *shadow*, and *size* as fixed factors and *name* as random factor. Only the depth modes: 'sinking background', 'floating icon', and 'window effect' are taken into account.

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Depth mode	3.467	2	1.734	4.777	.009
Background	8.608	1	8.608	23.717	.000
Shadow	1.185	1	1.185	3.266	.071
Size	.160	1	.160	.442	.507
Name	83.864	18	4.659	12.838	.000
Depth mode x Background	.385	2	.192	.530	.589
Background x Shadow	.308	1	.308	.849	.358
Background x Size	.058	1	.058	.160	.689
Depth mode x Shadow	.585	2	.292	.805	.448
Depth mode x Size	.298	2	.149	.411	.663
Shadow x Size	.149	1	.149	.411	.522
Error	153.517	423	.363[b]		