MASTER

Easy gesture based 3DTV interaction
the influence of stereoscopy, motion, and occlusion

Duijx, B.J.

Award date:
2009

Easy Gesture Based 3DTV Interaction:

The Influence of Stereoscopy, Motion, and Occlusion

AUTHOR

Name, initials, prefixes: Duijx, B.J. (Berrie)
Identity number: 0578588

SUPERVISORS

Name 1st supervisor: IJsselsteijn, W.A. (Wijnand)
Name 2nd supervisor: Kort, Y.A.W. de (Yvonne)
Name 3rd supervisor: Beurden, M.H.P.H. van (Maurice)
Faculty/university: Eindhoven University of Technology

Name 3rd supervisor: Ding, X. (Xinrong)
Name 4th supervisor: Quaedvlieg, C. (Christian)
Company/institute: Philips Consumer Lifestyle Innovation Lab.
City: Eindhoven

TIME PERIOD

Start date: October 1st 2007

PROPOSED VERSION

Filing date of first version: December 18th 2008
Filing date of this version: May 7th 2008
Version number: 1

Abstract

The effects of binocular depth information (stereoscopy), motion, and occlusion of a three-dimensional user interface (3DUI) displayed on a three-dimensional television (3DTV), were measured on the perceived 3D effect, willingness to use gestures as a natural interaction method, and on interaction device preference. Forty people participated in the experiment, in which they were interviewed and had to rate different interfaces. During the experiment reactions, comments and behavior were recorded. Significant effects were found on all three variables, stereoscopy showing the strongest effects, followed by motion, and lastly occlusion. People more often chose gestures as interaction method, as opposed to the more traditional interaction devices when the interface consisted of binocular depth cues. Furthermore, people perceive the interface as more graspable compared to an interface without binocular depth cues.

"The bottleneck in improving the usefulness of interactive systems increasingly lies not in performing the processing task itself but in communicating requests and results between the system and its user. The best leverage for progress in this area therefore now lies at the user interface, rather than the system internals. Faster, more natural, and more convenient means for users and computers to exchange information are needed. On the user's side, interactive system technology is constrained by the nature of human communication organs and abilities; on the computer side, it is constrained only by input/output devices and methods that we can invent. The challenge before us is to design new devices and types of dialogues that better fit and exploit the communication-relevant characteristics of humans.

Faster, more natural — and particularly less sequential, more parallel — modes of user-computer communication will help remove this bottleneck."

(Jacob, Leggett, Myers, & Pausch, 1993)

Index

Abbreviation

ABBREVIATIONS

3D - three-dimensional

3DTV - three-dimensional television

3DUI - three-dimensional user interface

DVB - digital video broadcast

DVR - digital video recorder

GUI - graphical user interface

HD - high definition

HDTV - high-definition television

IPTV - Internet protocol TV

LCD - liquid crystal display

LED – light emitting diodes

MC - media centre

PC - personal computer

RC – remote control

SOE - spatial operating environment

TUI - tangible user interface

TV - television

VOD - video-on-demand

WIMP - windows, icons, menus, and pointers

ZUI - zoomable user interface

1. Introduction

This study was conducted at the Philips Consumer Lifestyle Innovation Lab as part of the "Easy TV interactions" project, part of the European ITEA2 "Easy interactions" project. The goal of these projects was to aim at the design, prototype and validation of gesture-based 3DTV-control making use of virtual objects displayed on a 3DTV. Partner in the "Easy TV interactions" project was Philips 3D Solutions. Their input into the project was the 42-inch 3D WOWvx display, an autostereoscopic lenticular three-dimensional television (3DTV).

Figure 1: **3DTVs are expected to bring new viewing experiences to the living room.**

The world of TV is changing; three-dimensional televisions (3DTVs) are the latest development in display technology and are expected to bring a new viewing experience in the market of televised entertainment (Fehn, 2006). Some regard them as the logical step following high-definition television (HDTV) (Meesters, IJsselsteijn, & Seuntiëns, 2004).

The 3DTV has the possibility to create a three-dimensional user interface (3DUI), and because of this, people could have the feeling that such an interface is floating in front of the TV. As a result, they might feel as if the interface is graspable or tangible, and therefore, that gesture interaction is a logic or natural way of interacting with the 3DTV.

The goal of this study is to learn more about how we can interact with a 3DTV in an easy and natural way, such that everybody can understand the system and intuitively interact with it.

Our aim was to find out if a 3DTV enhances the willingness to use gesture interaction as a natural way of interacting with the TV in the living room.

Therefore, this study will start with a literature study, Chapter two, discussing the possible impact a 3DUI on a 3DTV could have on traditional interaction methods in the living room. It will discuss the changes in the living room cause by the changing world of TV, explaining our claim that the TV will be, even more than before, the center point of our living room. Because current input devices seem less suited for the future changes to come, new and more natural ways of interaction are discussed with a focus on gesture interaction. Furthermore, the 3DTV will be discussed in detail, covering different types of displays and explaining how they work. Finally, 3DUIs are discussed together with the impact those interfaces might have on interaction styles.

In addition, gain more information about people's thoughts about 3DTVs, 3DUIs, and gesture interaction in the living room; an exploratory pilot study was conducted, which is discussed in Chapter three. Now, with the literature, theories, and people's thoughts covered, Chapter four sums up the research questions and hypotheses for the main experiment, which will be discussed in Chapter five.

The main experiment will test a 3DTV with a 3DUI on perceived 3D effect, willingness to use gesture interaction, and on interaction device preference. The experiment will not only test the effects of binocular depth cues, but also test monocular depth cues such as motion and occlusion because the pilot study suggested that these cues could have substantial influence on the 3D effect and on the way people would be willing to interact with an interface.

Furthermore, participants from the main experiment were asked to visually show how they would interact with the 3DUI using gestures, to examine what the most commonly or naturally chosen gesture solutions would be. In addition, all participants were interviewed for their opinion on possible problems for a 3DTV in the living room. Problems such as unintended gesturing, who controls the interaction in a family setting, the physical effort of gesturing, and the lack of tactile feedback!

Finally, the conclusion in Chapter Six discusses the outcome of the study and argues why we think the 3DTV's binocular depth cues might just be the thing people need to cross the line and start interacting with the TV from a distance, using intuitive gestures.

2. Related Work: The Impact of a 3DTV in the Living Room

The world of televised home entertainment is changing. Continuous technological developments have changed the face of TV rapidly since the first television was patented back in 1884 (Okamura, 1994). In the early days, TVs used to be big and bulky, producing only black and white images. Nowadays, TVs have evolved into ultra flat, wide screen entertainment systems delivering crystal clear high-definition (HD) quality images.

In an attempt to create cinema like experiences in the comfort of a normal living room, TVs have become bigger, wider, and flatter. Currently, 150-inch TVs are available (Engadget, 2008), Philips holds the record for thinnest Liquid Crystal Display (LCD) TV with an astonishing 8-millimeter thickness (Engadget, 2008), and ultra widescreen is the on our doorsteps (See Figure 2).

Figure 2: **Philips' 56" conceptual cinema 21:9 HDTV**

And so the strive for new viewing experience in the market of televised entertainment continues, and in search of 3D cinema experience, 3DTV is ready to conquer the living room.

2.1 Living Room

With the 3DTV expected to come to the living room, this section will first discuss some typical living room settings and behavior in order to contemplate the differences the 3DTV might bring to the future living room.

2.1.1 Typical Living Room Setting

A typical living room can be described as a room in a private house or establishment where people can sit and talk and relax. A living room often is also called the family room, as families spend a lot of time there together. Families and television are practically inseparable. Families depend heavily on television for all sorts of information and entertainment. The average American watches more than 7 hours of TV each day (Readers Digest, 2009). Most often, the TV in the living room is surrounded by one or more couches or big chairs. The distance from the sitting position to the TV ranges from two to five meters, depending on the size of the room, the size of the TV, and on the position of the furniture. Although living rooms might look totally different (see Figure 3), fundamentally most look the same when merely looking at the TV to couch relation.

Figure 3: Two examples of typical living room environments.

2.1.2 Typical TV Viewing Behavior

Typical viewing behavior could be divided into passively or actively watching TV, or into more interactive forms of engagement such as gaming. Passive viewing can be described as someone whom only consumes information. This could often relate to lazy behavior (see Figure 4). With active viewing is usually meant that people have a retightened interest in the program content. They would probably display a more forward active posture, rather than a "lean back" relaxed posture.

Figure 4: **Examples of people lying in the sofa watching TV.**

Interactive forms of TV viewing behavior on the other hand could be described, as viewing that requires some form of interaction, such as playing video games for example. With the emergence of video games and gaming consoles, active TV viewing is becoming more regular. Especially with devices such as the Nintendo Wii, which lets you often participate with your whole body (see Figure 5).

Figure 5: **Examples of active physical gaming workouts on the Wii gaming console.**

In addition, some games require realistic replicas of existing devices or physical props to play the game. As can be seen in Figure 6, these devices could provide a feeling of 3D interaction because they feel and handle the same as their original counterparts.

Figure 6: **Different examples of gaming interface props such as guitars for the game Guitar Hero and light sabers, steering wheel, and a zapper gun for as Wii controlers.**

With this more active form of interacting with the TV, already, the setting of some living rooms have been seen to change. As can be seen in Figure 3, in the newer living room on the left, there is room to actively participate in those video games as discussed above. The older living room on the right has no physical space to actively participate. What happened is that the small living room table has vanishes from the living room setting; this makes room for active forms of TV interaction. Nevertheless, the distance from the couch towards the TV stays the same.

Now with the coming of the 3DTV to the living room, bringing true three-dimensional entertainment to the scene, it could be argued that these forms of active interaction, active 3D interaction would only increase. And as a result, having more influence on the living room setting and on the active way of interacting.

2.1.3 Integration of Functionalities in the Living Room

Already, typical TV viewing behavior is arguably changing because of more interactive forms of TV engagement, but also because the living room is becoming more digitalized. Different devices are being connected to each other, and most often have to be operated or controlled through the TV.

Wireless connectivity enables the TV to be connected to different devices from everywhere in the house without the need for cables, (i.e. to stream HD content (Engadget HD, 2008)). Central data servers can be installed outside the living room and serve as media centers, giving more freedom to the placement of the TV. In that case, controlling those devices will have to be done though the TV.

Figure 7: **An example of a Philips CitizenM hotel room, where you can customize the TV, ambient lighting and mood, temperature, and music preferences through the same remote control.**

But apart from media devices, home automation or domotics can also be seen as another form of integrating functionality into the TV. For example, Philips has demonstrated hotel TV concepts (Philips News Center, 2008), centrally combining the TV, radio, lights, and more to create an ambient experience in the hotel room (see Figure 7). Centrally controlling the ambient experience of the living room through the controls of the TV is something to expect from the future.

This makes the TV more and more serve as a central control point of technology in the living room. The increase of TV functionalities makes it harder to operate these devices through traditional input devices such as remote controls and calls out for new and more natural ways of interaction.

Now with the arrival of the 3DTV, menus and interfaces could be designed to make use of the binocular depth cue features of the 3DTV, making the interface appear in stereoscopic depth. These 3DUIs could bring an extra dimension to the way people are used to interacting with menus.

2.2 Interaction Method

All these changes call for new and more natural ways of interaction, as current input devices seem less suited for the future to come. For a normal household it is not uncommon to have four or more remote controls to control all the different devices in the living room. Programmable or universal controls are available but often make the interaction only more complicated, and to handle a keyboard and mouse while sitting on the couch is surely not ideal. As for gaming controllers, their complexity increased with some having two analog joysticks, a directional pad, and 10 buttons, making games more difficult to learn and master (Joseph & LaViola, 2008). Overall, controllers can get lost, batteries can run dead, or the consumer just does not know where to start anymore. It is important to study how we can keep the interaction easy and natural, in a way that everybody can understand the system and intuitively interact with it. Using gestures as interaction method could solve some of these problems.

2.2.1 The definition of a gesture

Gestures are defined as physical movements of the hand, arm, face, or body with the intention to convey information or meaning. Biologists define the term 'gesture' as "the notion to embrace all kinds of instances where an individual engages in movements, whose communicative intent is paramount, manifest, and openly acknowledged" (Nespoulous, Perron, & Lecours, 1986). Gestures can be divided into two categories; gestures associated with speech or independent of speech. Those associated with speech are referred to as gesticulation, while gestures that are function independently of speech are referred to as autonomous or functional gestures.

Cadoz divided the class of functional gestures into three groups: semiotic, ergotic and epistemic (2004). Ergotic gestures are those that manipulate physical objects in the world, while semiotic gestures communicate information (Rime, & Schiaratura, 1991). 'Thumbs up' is an example of a semiotic gesture. Finally, epistemic gestures are exploratory motions that gather information from the environment (Kirsh, 1994; Kirsh & Maglio, 1994).

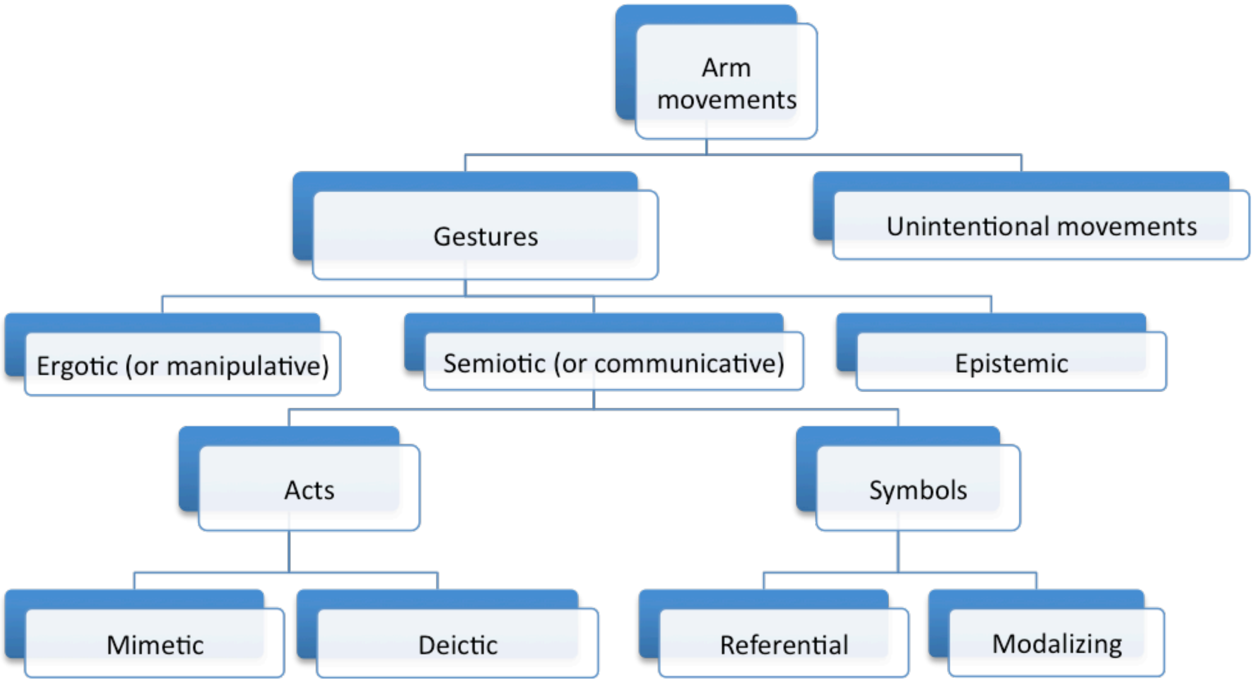


Figure 8: **Arm movement categorization**

Semiotic or communicative gestures can be further classified as being iconic, metaphoric, deictic, or beat-like (McNeill, 1992). Iconic gestures are representatives of an action, object or event. Metaphoric gestures depict a common metaphor rather than depicting the event or object directly. Deictic gestures are pointing gestures that are used to indicate people, objects, or directions. And beat-like gestures are small, emphatic gestures generally performed with the hand or the head.

Semiotic gestures could also be summarized in either acts or symbols. Acts are gestures that are directly related to the interpretation of the movement itself, classified as either mimetic, which imitates some actions, or deictic, which are a pointing acts. Symbols are gestures have a more linguistic role: they symbolize some referential action or are used as a modalizer.

2.2.2 Natural and Intuitive Input Devices

There appears to be a trend towards more natural and intuitive input devices (Subramanian & IJsselsteijn, 2000). Already the game industry is providing devices that enable more realistic gameplay using replicas of existing devices or physical props (for example the guitar in Guitar Hero, see Figure 6). These devices don't necessarily provide 3D interaction, but other approaches do. For example, a Wii controller (Wiimote) lets you interact and control elements of the 3D gaming world through motion sensing. Apart from the fact that it is still a physical device, another disadvantages is that because of small vibrations and unreliable sensors, the Wiimote often is imprecise with small interactions.

Figure 9: **Example of the use of the Sony EyeToy**

Alternatively, there are vision-based sensing solutions, used for example with the Sony EyeToy. The Sony EyeToy connects the TV to a small camera and reads the gestures of the people in front of the TV, integrating them in to video games (see Figure 9). Such a solution has the advantage that you do not need external devices to create the interaction. Cameras record the users body and gestures, from which the spatial input is interpreted.

Systems as mentioned above, could be enhanced with stereo cameras creating an even more accurate spatial 3D input of the interaction, which could enhance the way people could naturally and intuitively interact with a 3DTV.

2.2.3 3D Gesture Recognition

Currently there are a couple of companies working on 3D gesture recognition, which could be used in different applications such as gaming, web-conferencing, home fitness, or for controlling the digital home. Companies such as Canesta, Soft Kinetic, 3DV, and Prime Sense are all working on unique imaging technology capable of capturing the depth dimension of objects in real time. These technologies often enable sensing motion and recognizing shape within a dynamically defined three-dimensional space (see Figure 10). In addition, some systems are complemented by the ability to combine your own 3D image inside a scene in real-time.

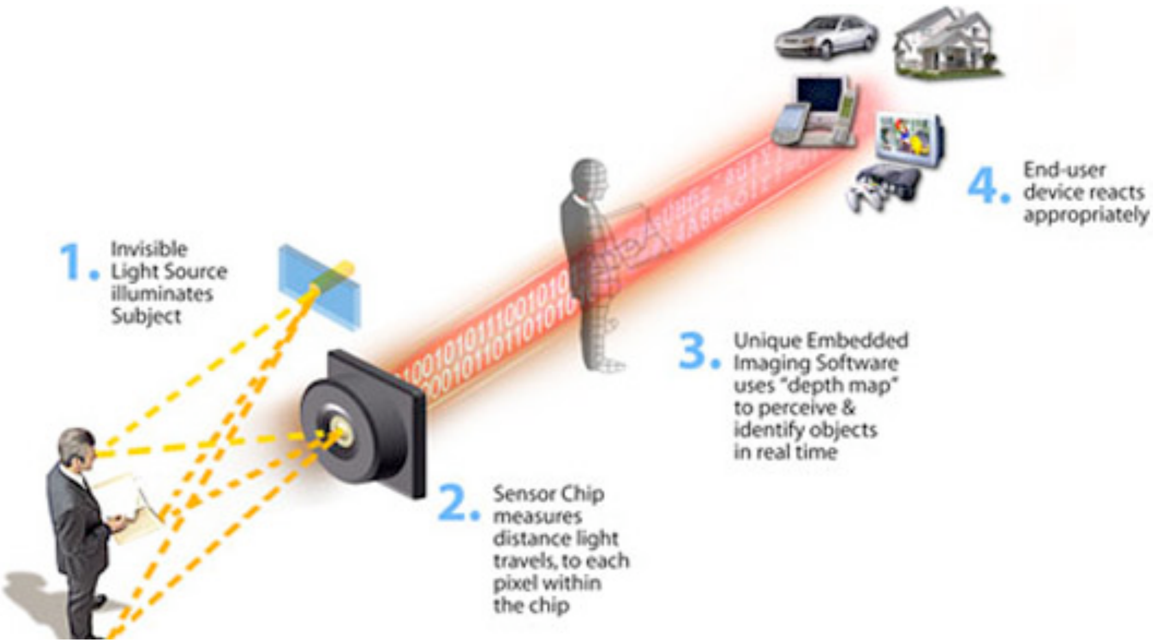


Figure 10: **A simple overview of how 3D capture devices might work.**

Take gaming for example; the success of the Nintendo Wii clearly is due to in the way of interacting with the game. With 3D recognition, no controller would be needed and the whole body could act as a natural interaction device (see Figure 11). Supposedly, Microsoft is also working on a motion-sensing add-on for their Xbox 360 game console that could detect full body movements (Engadget, 2009). For example, in fighting games users would be able to kick, punch, duck, dodge, dive, jump, and so forth with their whole body.

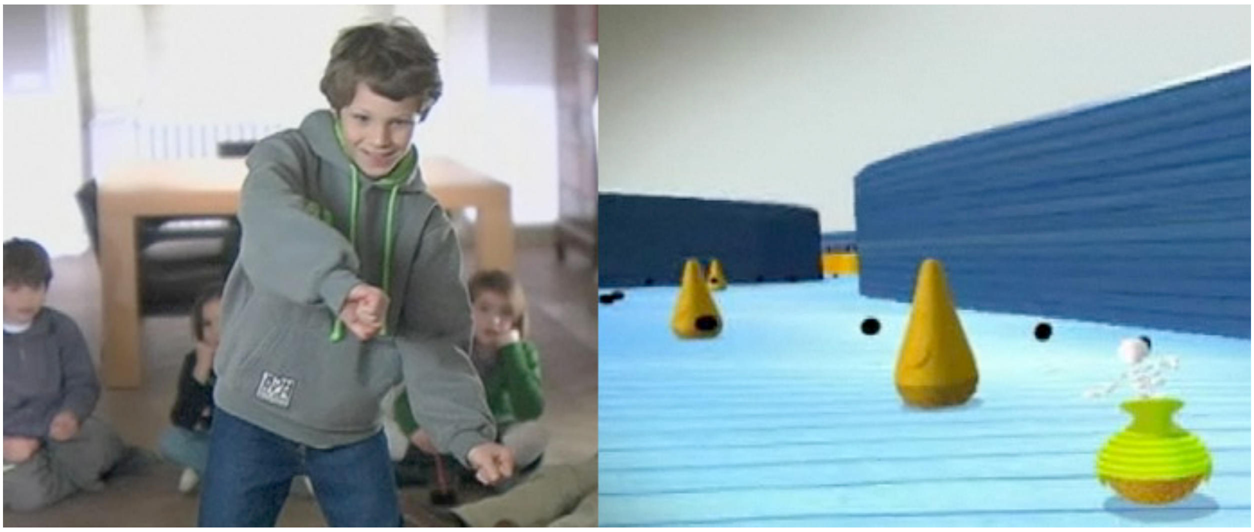


Figure 11: **An example of a Soft Kinetic game application.**

The way video games are being played is currently in an early stage of a revolution (Joseph & LaViola, 2008), and it can be argued that this will spread throughout other forms of media interaction. Using vision-based sensing with gesture as an interaction method could make interacting with the 3DTV feel more natural and intuitive. However, classifications of the vast range of gesture-based research have concluded that much has been done in theory, but little has been applied in practice (Karam & Schraefel, 2005).

Nevertheless, gestures seem suited for these tasks, “the past 40 years of computer research that includes gesture as an interaction technique has demonstrated that gestures are a natural, novel and improved mode of interacting with existing and novel interfaces” (Karam & Schraefel, 2005 pp. 38). Research has shown that one’s bare hand is more practical than traditional input devices (Hardenberg & Brard, 2001). Arguably, gesture based interaction could be more practical than interaction with a remote control, when interacting with a 3DTV in a living room situation.

It has also been shown that deictic gestures, which involve pointing to establish the identity or spatial location of an object within the context of the application, are potentially useful for all kinds of selections in human computer interaction (Bolt, 1980; Lenman, 2002), and arguable, could also be useful for human 3DTV interaction.

Controlling the PC or TV through gestures could replace the keyboard or remote control. Recently, Hitachi and Canesta demonstrated a gesture-controlled Hitachi television at the Winter Consumer Electronics Show (see figure 12). "The excitement among attendees and media observers was widespread, and there was much talk about how this would change the digital living room," (Canesta, 2009).

Figure 12: **The Hitachi and Canesta gesture-controlled TV prototype.**

Already, machine vision-based techniques, such as pointing and gestures, have been developed to enhance traditional human-computer interaction (Hopf, 2006), and hence, could benefit future users in human-TV or human-3DTV interaction. In addition, combining a 3DTV with gestures as interaction method has the advantage that people no longer need external devices to interact with their TV.

2.3 Three-Dimensional Television

The next big thing in home entertainment will be the introduction of 3DTV into the living room. Three-dimensional or stereoscopic televisions have been around since the 1920s, but have only been presented to a large audience in an experiment in the 1980s (Sand, 1992). However, poor image quality and unwanted side effects limited the successful introduction to the mass consumer market. Nevertheless, recent developments in display technology are expected to bring the 3DTV to a new viewing experience in the market of televised entertainment (Fehn, 2006).

Figure 13: **The Philips 3D Solutions' WOWvx autostereoscopic lenticular 3DTV.**

The following section will explain concepts needed to understand this technology, followed by a description of the different types of 3D displays, and the final part will discuss the importance of 3D content.

2.3.1 Depth Perception

Viewing our surroundings is obtained through our visual system and the forms of information used for this purpose are often referred to as "depth cues". These cues are often grouped based on common features (Table 1).

Depth cues can be classified in three ways; firstly by Retinal Image or Oculomotor. This means that depth cues can be identified merely by the image on the eye's retina, or that the brain needs extra information from the eye's muscles to identified the depth cue, the later is called oculomotor depth cues. The second way to classified depth cues is by Pictorial or Parallax. Pictorial cues are cues available in two-dimensional (2D) static projections such as photographs, televisions, computer screens, or cinemas. Parallax cues require a different viewpoint in either time or space. A final way to classify depth cues is by being Monocular or Binocular. The first are cues that can be identified with only one eye, the later require input from both eyes.

Table 1: **Classification of depth cues grouped to common features (Blundell & Schwarz, 2005 p77).**

	Classification Scheme		
Depth Cue	Retinal Image / Oculomotor	Parallax / Pictorial	Binocular / Monocular
Binocular parallax (stereopsis)	Retinal image	Parallax	Binocular
Motion parallax	Retinal image (+Oculomotor)	Parallax	Monocular
Linear perspective	Retinal image	Pictorial	Monocular
Occlusion (interposition)	Retinal image	Pictorial	Monocular
Familiar size	Retinal image	Pictorial	Monocular
Shading and shadows	Retinal image	Pictorial	Monocular
Aerial perspective	Retinal image	Pictorial	Monocular
Texture gradient	Retinal image	Pictorial	Monocular
Height in the visual field	Retinal image	Pictorial	Monocular
Accommodation	Oculomotor	-	Monocular
Convergence	Oculomotor	-	Binocular

For an extensive description of perception, vision and depth cues consult to chapter 2.6 the book "Creating 3-D Displays and Interaction Interfaces" by Blundell & Schwarz (2005). Figure 14 shows the Just-discriminable ordinal depth thresholds as a function of the logarithm of distance from the observer for most of the depth cues from Table 1. The next sections will describe some of those depth cues.

Binocular parallax (Stereopsis): Because each of our eyes is situated at different spatial positions, both see a slightly different view of a 3D scene. The retinal reflection of an object could be slightly offset in each eye. This disparity between corresponding points between the two reflections provides a powerful sensation of depth.

Movement or motion parallax: Parallax cues are also available from temporal changes in either the scene or the viewer's position. Motion changes of different features in a 3D scene can provide a strong sensation of depth.

*Occlusion (Interposition)****:*** When an object is partially covering another, it is believed that that object is in front of the other and hence closer.

Linear perspective: This cue arises when an object converges and creates a smaller retinal image, making the widest point closer. For example, the convergence of railway lines in the distance to a vanishing point.

Familiar size: The familiarity of objects can provide an absolute indication of their distance and hence help to calibrate the depth of features within a scene. Knowing real-world sizes of many objects help to interpret its content.

Aerial perspective (Atmospheric haze): Objects that take on a blue hue and appear less distinct or look more washed out, are judged to lie at a greater distance. For example, distant mountains will have a slightly bluish hue because of the scattering of light from small particles in the atmosphere.

Texture gradients: A texture surface can provide another depth cue when it includes a repeating pattern or variations in intensity or color. Those elements in a texture are perceived larger in the retinal image when they are closest to the observer. The image in the retina decreases gradually for the parts of the texture that are further away.

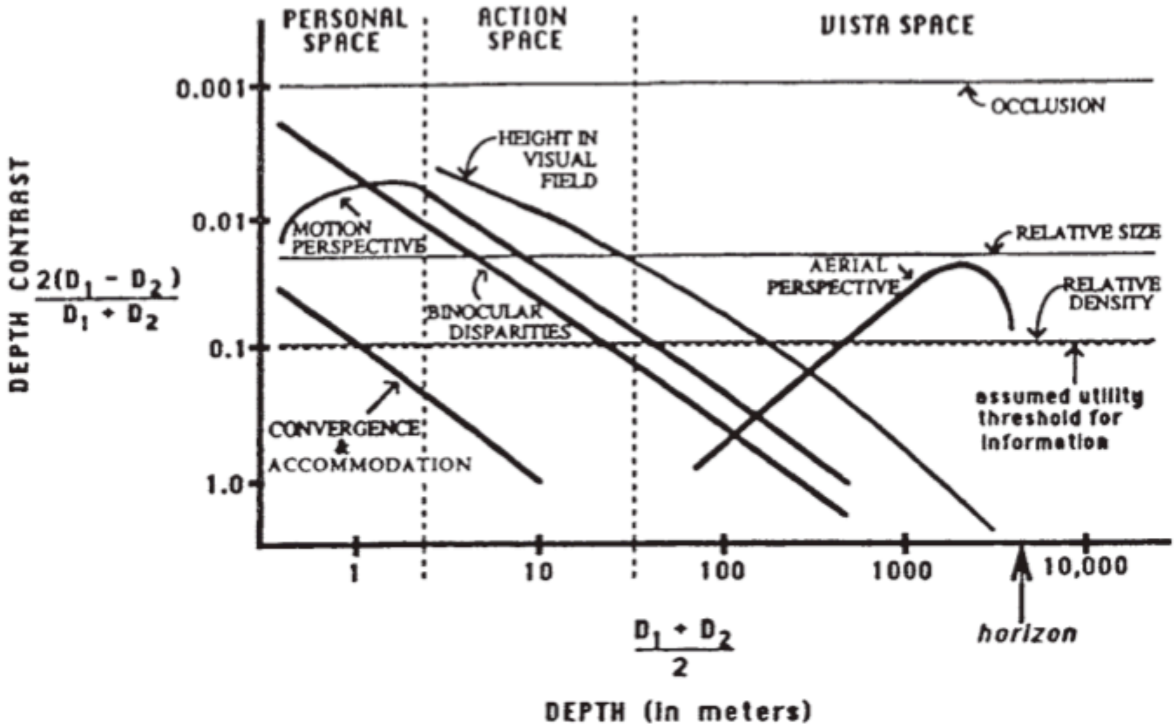


Figure 14: **Just-discriminable ordinal depth thresholds as a function of the logarithm of distance from the observer, from 0.5 to 10,000m, for nine sources of information about layout.** (Cutting, 1997 p. 29).

Vergence: The process when our eyes center their gaze on the same object. To obtain depth information, the brain uses the muscular forces exerted on the eyes. As a depth cue, convergence is most effective for nearby objects.

Accommodation: The process when the shape of the lens of the eye alters in order to focus on a specific point. To obtain depth information, the brain uses the amount of the tension applied to these muscles. Like convergence, accommodation is most effective for nearby objects.

2.3.2 3D-Displays

3D displays can be divided into two different types, stereoscopic 3D displays and autostereoscopic 3D displays. The main difference is that stereoscopic 3D displays need special goggles to see the 3D effect, whilst autostereoscopic 3D displays do not need additional goggles; the 3D effect can be seen with the bare eyes.

Stereoscopic 3D displays

A stereoscopic 3D display can be divided into three different systems (Dodgson, 2005), anaglyph, polarized, or a shutter system, and all systems need specific goggles to see the 3D effect on the TV screen. In addition, head-mounted displays could be added to this category.

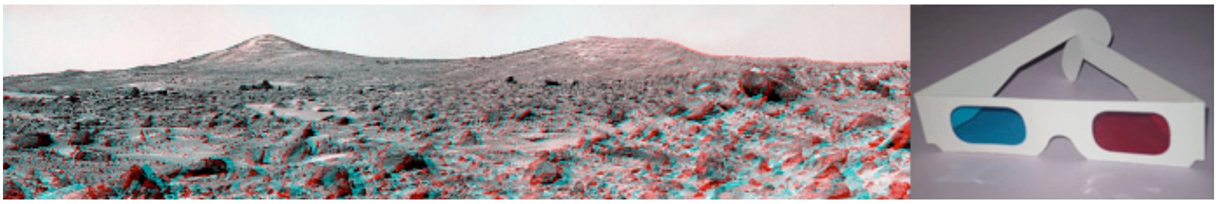


Figure 15: **Example of a stereoscopic anaglyphic picture and the colored glasses.**

An anaglyphic system (Wikipedia, 2008) makes use of a normal colored display combined with colored glasses. Most commonly known are the goggles with red and green (cyan) colored glasses (see Figure 15). The content is made out of two colored layers that are superimposed onto each other, but offset with respect to each other to produce the depth effect. When the content contained in the two different colored images, one for each eye, are viewed through the special goggles, the brain fuses this into a three dimensional scene.

A polarized system makes use of a standard display with a polarized filter combined with polarized goggles. Two different streams, one for each eye, are interlaced using a polarized filter. Using polarized goggles, this facilitates the eyes to see two different images. Again, the visual cortex of the brain fuses these two images into one three dimensional scene. Commercial systems are available such as the StereoMirror from Planar (Planar Systems Inc., 2009), which uses two displays and a polarized mirror to achieve the 3D result (see Figure 16). Other, more recent products use only one monitor with a polarized sheet directly on the display to achieve the same 3D effect. For example, Miracube 3D display from Pavonine (Engadget, 2008; Pavonine, 2009).

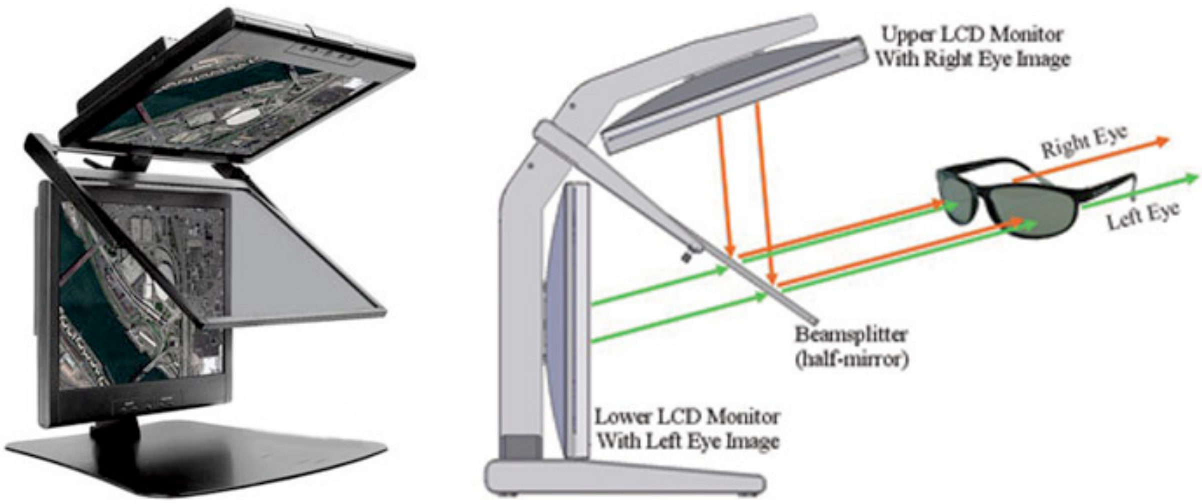


Figure 16: **Example and workings of the StereoMirror by Planar.**

A shutter system makes use of a normal display in combination with shutter glasses. The two streams are swapped frame after frame, and at the same time, the goggle shuts the left or the right glass. In this way, both eyes see a different stream resulting in a three dimensional image. Samsung is currently shipping its "world's first 3D-ready flat-panel HDTV" (Engadget, 2008) based on shutter glass technology from DDD TriDef (McAllister, 1993).

Head-mounted displays are goggles on which two small displays are attached. The two displays deliver different video streams directly to the eyes, which results in a three dimensional representation (for examples see Figure 17).

Figure 17: **Examples of head-mounted displays.**

AUTOSTEREOSCOPIC 3D DISPLAYS

Autostereoscopic 3D displays can be divided into three different systems (Dodgson, 2005), two view displays, head-tracked displays (two view), or multiview displays. Autostereoscopic displays provide the 3D image without the viewer needing to put on any special viewing gear (Benton, 2001). Although they do not make use of a TV like display, volumetric and holographic displays could also be added to this category because they also create a 3D effect without the need for special goggles.

To make two-view autostereoscopic displays, researchers have been using either parallax barrier or lenticular sheet technology for more than a century (Benton, 2001; McAllister, 1993; Okoshi, 1976). The parallax barrier technology makes use of a barrier mask in front of the pixel raster. In this way, each eye sees light from only every second pixel column of the screen. Lenticular sheet technology makes use of an array of cylindrical lens-lets placed in front of the pixel raster. In this way, light is directed from the adjacent pixel columns to different viewing slots at the ideal viewing distance. As a result, each of the viewer’s eyes sees light from only another pixel column, depending on the number of views the display has.

When standing in the correct position and at the ideal distance of two-view autostereoscopic displays, the viewer will perceive a stereoscopic image (Dodgson, 1996). However, the disadvantage of two-view autostereoscopic displays is that there is a 50 percent chance to see an incorrect, pseudoscopic image because the viewer is in the wrong position. Furthermore, the viewer must stay fairly still to stay in the correct viewing position.

A two view head-tracked display knows the position of the viewer's head. This prevents pseudoscopic viewing because the system can display the right and left image into the appropriate zones. However, problems with head-tracked display can occur when the viewers' eyes separation differs significantly from what the systems expects. A solution could be to use eye tracking instead of mere head tracking.

Finally, multiview autostereoscopic displays hold the advantage that viewers perceive a stereoscopic image wherever they are within the viewing zone. This type of display can accommodate multiple viewers and each will see a 3D image from their own point of view (Dodgson, 1996).

VOLUMETRIC OR HOLOGRAPHIC DISPLAYS

Another way to create a three dimensional image is to use volumetric or holographic displays. These differ from the autostereoscopic examples mentioned above because do not use a TV-like flat screen. There are different types of volumetric and holographic 3D displays (see Figure 18). Swept-volume displays make use of illuminating rapidly moving parts such as varifocal mirrors or rotating screens. Static-volume displays do not make use of moving part but use for example fixed leds, or occasionally use laser light to encourage visible radiation in a solid, liquid, or gas.

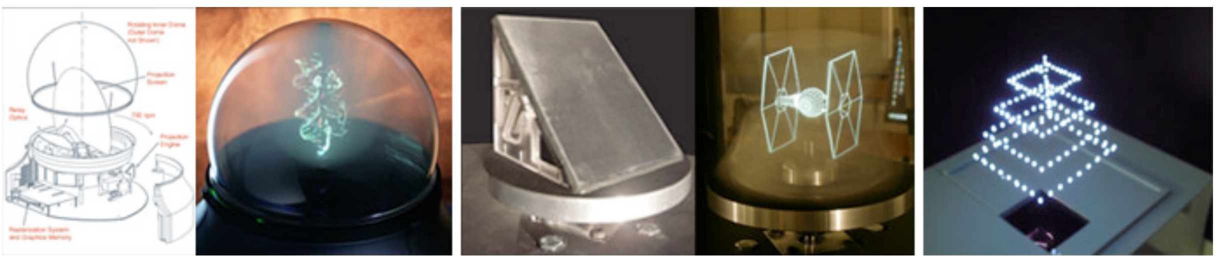


Figure 18: **Left: Perspecta Spatial 3D, a swept surface rotating screen Volumetric display by Actuality systems. Centre: Volumetric display using a spinning mirror. Right: Static-volume volumetric display using infrared laser pulses to create balls of glowing plasma.**

2.3.3 Content in the Third Dimension

The success of the 3DTV will depend largely on the availability of 3D content. Content for the 3DTV can be divided broadly into three categories, movies, TV broadcasts, and games.

The driving power behind 3D movies will be Hollywood, and has produced at least 30 or more 3D movies (Alternative Film Guide, 2008) for the year 2008. They also announced deals to convert as many as 10,000 more theatre screens for the digital technology needed to accommodate the resurgent 3D format. "Access Integrated Technologies Inc. (AccessIT) said it had reached agreements with four studios — Disney, Twentieth Century Fox, Paramount and Universal Pictures — to finance and equip the screens in the United States and Canada during the next three years" (Alternative Film Guide, 2008). The customer demand for cinematic 3D experiences at home will pull the 3DTV into the living room.

Apart from movies, broadcasting will also play a major role in the success of 3DTVs. Already depth-image-based-rendering (DIBR) techniques (Mark, 1999; McMillan, 1997) are available with the advantage that they are backwards compatible with today's 2D digital TV (Fehn, 2003; 2004). In addition, the acceptance and widespread introduction of digital broadcasting makes the transmission of a stereoscopic signal increasingly feasible (Meesters, IJsselsteijn, & Seuntiëns, 2004). Experiments and trials have been conducted to broadcast live sport event in 3D, such as the 2002 FIFA World Cup (Hur, Ahn, & Ahn, 2002) and the Six Nations Rugby union Championship (BBC News, 2008). Even entire concerts (U23D, 2008) are being produced in 3D.

And finally the games industry will supply content for the 3DTV. Although concrete comments about stereoscopic 3D gaming for the consoles are scarce, some companies are speaking. Blitz Games Studios has both a software engine that enables a PlayStation 3 or Xbox 360 to output stereoscopic 3D content, and is working on a native stereoscopic 3D game title (GameCyte, 2009). They claim that their technology will give games developers the opportunity to recreate a full 3D HD experience "with the 3D images 'popping' out of the 3DTV screens and into living rooms for the first time" (TechRadar.com, 2008).

A particular interesting feature that might be included with 3D games will be that the user has to interact with the game or the menu in 3D as well, opposed to watching a movie or a TV broadcast which requires no interaction. With 3D games on the 3DTV, a new door opens towards 3D interaction and 3D interfaces.

2.4 Interface

With the advent of the 3DTV, one could question the traditional 2D user interface. For the last 30 years, people have been interacting with PCs and media devices through a graphical user interface (GUI) using the windows, icons, menus, and pointers (WIMP) paradigm. The most familiar GUI desktop must be the Windows operating series (95, 98, NT, ME, 2000, & XP), yet Mac OS and some Linux versions also make use of the same paradigm.

Recent trends in desktop technology are to integrate 3D effects into the interaction. However, most of these 3D effects, often called 2.5D effects, are achieved by using monocular cues such as cast shadows, transparency, occlusion, or motion. For example, the latest Windows version, Vista, and Mac OS 10 are both stacked with 3D effects. Nevertheless, they both still have the same 2D point & click interface.

Research is being conducted to create virtual desktops that behave more physically realistically or to resemble the look and feel of the real world. Bumptop (Agarawala & Balakrishnan, 2006) for example, is a virtual desktop that uses piles as the fundamental organizational metaphor, and physics simulation affording casual, potentially more realistic interaction (see Figure 19).

Figure 19: An example of a **BumpTop computer desktop, showing piles and stack of documents making use of 3D presentation.**

A logic step toward creating an appropriate interface for the 3DTV would be to realistically make use of the third dimension. In addition, realistic interactions are needed to stimulate more natural and intuitive forms of interaction. Nevertheless, people have been stuck on the WIMP paradigm for the last 30 years and might have trouble switching to another type of interface, let alone understanding it.

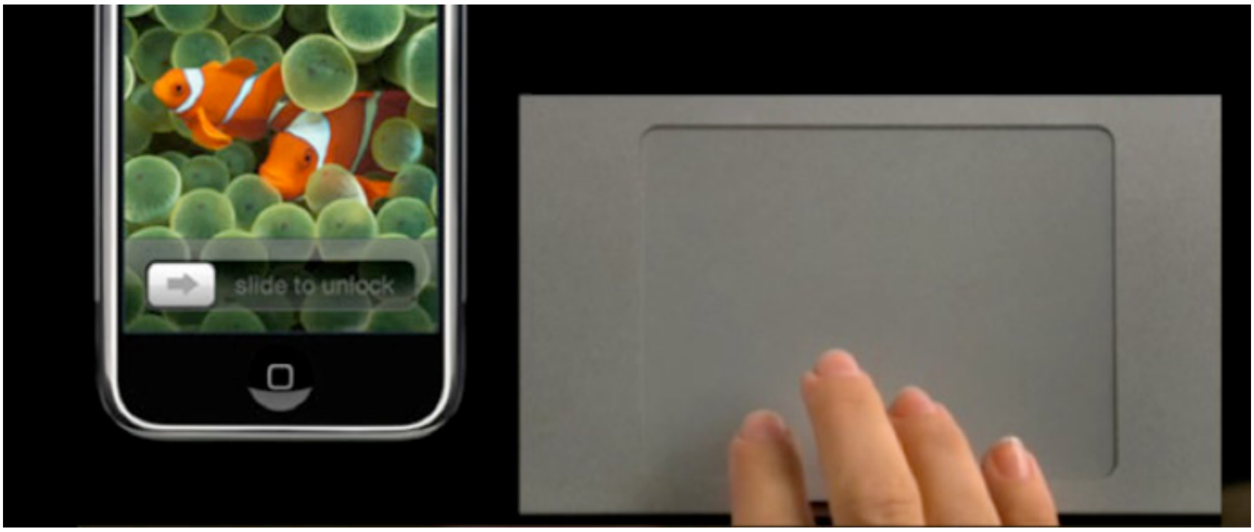


Figure 20: **Apple's multi finger gesture recognition solutions, the iPhone (Left) and the Macbook touchpad (right).**

2.4.1 Zoomable User Interface (ZUI)

Another ongoing trend in interface design are zoomable user interfaces (ZUIs). Basically, these are graphical environments where users can change the scale of the viewed area to see more or less detail. Most of the time, these actions rely on two handed or multi-touch interactions. One of the best examples is the success of the Apple iPhone (Apple Inc, 2008), which relies heavily on ZUI techniques (Elgan, 2007). Using two fingered interactions techniques, the iPhone can zoom in and out of documents, pictures, or websites to see larges amounts of data on only a small screen (see Figure 20).

Notebooks such as the Apple Macbook are equipped with multi finger touchpad gesture recognition (Engadget, 2008). Furthermore, desktops are also stepping into touch screen interaction for the consumer market. For example, the HP TouchSmart (HP, 2008), or Microsoft newly announced Windows 7 Operating System (Telegraph.co.uk, 2008) will be fully touch and gesture based.

Figure 21: **Microsoft Surface tabletop computing combines direct interaction, multi-touch contact, multi user experience, and object recognition.**

Current research often combines ZUIs with multi-touch surfaces. Multi-touch interfaces have the ability to directly touch and manipulate data on the screen without using any intermediary devices and have a very strong appeal to users (Benko, Wilson, & Baudish, 2006). For example (see Figure 21), Surface is a Windows Vista "coffee table" computer that features these multi-touch techniques (Microsoft Corporation, 2008). Surface needs no mouse or keyboard, and on-screen objects move like real 3-D objects and are manipulated directly by touching the screen (Elgan, 2007).

Other examples of multi-touch surface research are, the Sensetable (Patten, Ishii, Hines, & Pangaro, 2001), which explores multiple tangible objects interacting with a tabletop display. Tangible User Interface (TUI) research (Fitzmaurice, Ishii, & Buxton, 1995; Ishii& Ullmer, 1997), an essential component of Ubiquitous Computing and Augmented Reality research, has become increasingly widespread over the past 25 years (Blackwell, Fitzmaurice, Holmquist, Ishii, & Ullmer, 2007). Touchlight (Wilson, 2004), which explores multi-touch, gesture based, interaction on a semi transparent sheet of acrylic plastic. And PlayAnywhere (Wilson, 2005), which is a front-projected computer vision-based interactive table system.

One profound advantage of multi-touch interface design is that the interface could disappear; some researchers describe multi-touch interfaces as being completely intuitive, an interface without instruction manual, “it just does what you expect” (Han, 2006).

These same attributes, completely intuitive and natural, are attributes to keep in mind when creating a 3DUI for the 3DTV.

2.4.2 Three-dimension User Interface (3DUI)

A good 3DUI should focus on the means by which a human interacts with a virtual environment (VE) in a 3D spatial context (Bowman, Kruijff, Joseph, LaViola, & Poupyrev, 2004), and not merely on adding a third dimension to an existing interface. In other words, 3DUIs should include media for 3D representation of a system state, and media for 3D user input.

Figure 22: **The IllusionHole is a system with an interactive display that allows three or more moving observers to simultaneously observe stereoscopic image pairs from their own viewpoint.**

Now, with the ability of the 3DTV to add binocular cues to the perception of the interface, one could create a real three-dimensional user interface. A 3DUI could look more realistic and feel more tangible summoning a more intuitive way of interacting. Already some research has been combining tabletop computing and 3DTV. For example (see Figure 22), the IllusionHole is an interactive stereoscopic display system where users could simultaneously observe adequate stereoscopic images of data volumes for medical use (Kitamura, Konishi, Masaki, & Kishino, 2001). However, the input of this system is still conducted through traditional input devices such as a mouse.

In search for a suited 3DUI for the 3DTV, one should focus on the means by which a human interacts with a 3DTV in a 3D spatial context. As mentioned above, 3DUIs consist of several building blocks including virtual environment (VE) devices and 3D interaction techniques (3DITs) (Bowman, 2004). The next section will discuss these in further detail.

VIRTUAL ENVIRONMENT (VE)

Virtual environment is a computer-simulated environment from the field of virtual reality technology, in which a user can interact with a simulation of the real world or an imaginary world.

Virtual environments have several applications for work in the real world including simulation, medical, training, scientific visualization, designing and prototyping, collaboration, entertainment, and there are many specific examples of the benefits of working with VEs in these applications.

Immersive VEs are effective in the psychiatric treatment of phobias such as arachnophobia (Carlin , 1997), fear of heights (Hodges, 1995), and other mental ailments such as post-traumatic stress disorder (Hodges, 1995). Medical professionals enjoy the benefits of VEs with remote surgical training (Cruz-Neira , 1999). Immersive VR (Forsberg, 2000) and collaborative VEs (Kitamura, 2007) are used for visualization in medical applications. Scientists and engineers can use 3D graphics paired with 3D user interfaces to interactively obtain insight from complex data (Bryson, 1996).

Figure 23: **Military personnel using a VR parashute trainer.**

Pilots, military personnel (see Figure 23), firefighters, and such can also be trained in VEs allowing them to build experience while remaining in the safety of a VE (Cruz-Neira, 1999; Bliss, 1997). Furthermore, in the design domain, there exist VE tools for 3D modeling (Butterworth, 1992) and CAD (Kasik, 2002; Liang, 1994), designing and prototyping architecture (Liang, 1994) and automobiles (Cruz-Neira, 1999) where the immersive nature of VEs allow designers to more effectively visualize their designs. And finally, VEs are used for entertainment and gaming (Pausch, 1996).

As can be seen above, virtual environments are relevant to various sectors in society with so many different applications covering so many disciplines, and therefore, are valuable to many organizations in research, academia, industry, medicine, entertainment, the military, and more.

Virtual environments are very different from traditional desktop systems and as such have different requirements in user interface design. Traditional interfaces and interaction techniques are meant for the two dimensional realm while virtual environments use three-dimensional spatial contexts.

The 3DTV also uses three-dimensional spatial contexts, and therefore, also requires a new set of input devices, and interaction techniques to successfully interact with the new virtual environment.

Three-dimensional Interaction Technique (3DIT)

Three-dimensional interaction techniques are methods that are used in order to perform different tasks in a 3D space. 3DIT are also known in literature as 3D Interaction Component (3DIC) (Figueroa, Dachselt, & Lindt, 2006), 3D gadget (Schönhage & Eliëns, 1999), or 3D widget (Brookshire Conner, Snibbe, Herndon, Robbins et al., 1992).

The book on 3D user interfaces by Bowman et al. (2004) gives a recent overview of the numerous 3D interaction techniques that have been developed by various researchers during the past decade. Here the 3D interaction techniques have been classified using the following categories: manipulation and selection, navigation, symbolic input, and system control.

Manipulation and selection enable the user to interact with virtual objects. Navigation provides the user with information regarding location and movement, and therefore, can be divided into wayfinding and travel. Symbolic input allows the user to enter for example text, whereas system control allows a user to issue commands to the application.

The interaction techniques used in the 3DUI of a 3DTV should feel natural and intuitive and should provide more benefits opposed to traditional interaction devices used in the living room.

Figure 24: **The g-speak input framework from Oblong allows direct, either-handed, multi-person manipulation of any pixel on any screen you can see. Pointing is millimeter-accurate from across the room. Hand pose, angle, and position in space are all available at 100 hertz, with no perceptible latency and to sub-millimeter precision.**

Recently, a company called Oblong (Oblong, 2008) has created real-world implementation of the computer systems seen in Steven Spielberg´s Hollywood movie Minority Report. It is called g-speak (Engadget, 2008; Tweakers, 2008) and categorized as a Spatial Operating Environment (SOE). The g-speak system (see Figure 24) looks as a promising solution for natural 3D interaction techniques; however, it uses data gloves for gesture recognition. This makes the system less natural than a bare-hand solution. In addition, if such a system could be enhanced with autostereoscopic 3D technology, adding binocular depth cues could make the system more intuitive or inviting to use gesture interaction.

Looking at the near future of human computer interaction, we expect large changes in the upcoming years, which are inline with different media expectations (Cnet Australia, 2008).

2.5 Rationale

With a 3DUI displayed on a 3DTV, people could have the feeling that the interface or menu is floating in front of the 3DTV in a sort of spatial contexts. They might feel as if the interface is graspable or tangible, creating a spatial metaphor in front of the TV.

People might link the spatiality of gestures with the spatial metaphor the 3DUI creates, and as a result they might feel as if gesture interaction is a logic or natural way of interacting with a 3DTV. Gesture interaction would then serve as the 3D interaction technique. In this case the 3DUI would consist of a virtual environment (the 3D menu on the 3DTV), and of 3D interaction techniques (the 3D gesture interaction).

It would be interesting to find out if the 3D effect of a 3DUI on a 3DTV can be related to that tangible or graspable feeling. Therefore, research is needed to examine to what extent the 3DTV has influence on the perceived 3D effect of a 3DUI.

Furthermore, do people really feel as if gesture interaction would be a logical interaction method for such a system and would they feel as if the interaction is more natural and intuitive? Does a 3DUI have influence on the willingness to use gesture interaction, and again, what attributes of the 3DUI or 3DTV are responsible for these effects?

Finally, would such a system work in the typical living room environment? What problems could occur when people switch to these new interaction techniques and away from the traditional paradigm?

3. Pilot Study

Prior to the start of the main study, an exploratory pilot study was conducted with the goal to gather more global information about how people think about the combination of 3DTV, 3DUI, and gesture control in the living room.

More detailed information about the interview and the questionnaire, and the examples used in both, can be found in Appendix A. For a summary of interviews see Appendix B, and the raw data from the questionnaire can be found in Appendix C.

3.1 Method

Design

A semi-structured interview was held and a questionnaire was administered on the topic of 3DTV and 3DUIs. All interviews were recorded, and participants signed an informed consent to grant permission to use these recordings for further evaluations.

The interviews were held with the intention to ensure that the same general areas of information are collected from each interviewee, yet allowing a degree of freedom and adaptability in getting the information.

During the first part of the interview, participants were shown seven UI examples, and asked to give their opinion to different questions accordingly. In the second part, participants were shown seven 3DUI examples and are asked to fill in a questionnaire.

Participants

There were 15 participants (12 male and 3 females, ranging from 25-53 *M*: 36,33), all employees of the Philips Consumer Lifestyle Innovation Lab. All participants were familiar with the 3DTV in a sense that they had seen it at least once in a demo or product presentation. Participants were briefed about the purpose of the project at the beginning of the interview.

Apparatus

A 42inch Philips 3D Solution WOWvx 3DTV connected to a normal personal computer was used to show the 3D content. A notebook was used to show the normal UI screenshots, and recorded the interview using MP3 recording software. During the interview comments and notes were written down.

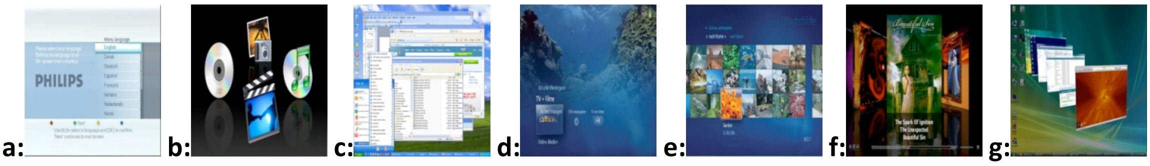


Figure 25: **Pictures of the seven UI examples that were used during the interview of the pilot study (a: Philips Easy Logic TV UI, b: Apple Front Row UI, c: Windows XP UI, d: Vista Media Centre UI, e: Media Centre Image Viewer UI, f: Apple Cover Flow UI, and g: Microsoft Vista UI).**

Seven UI examples were used (see Figure 25): a: Philips Easy Logic TV UI, b: Apple Front Row UI, c: Windows XP UI, d: Vista Media Centre UI, e: Media Centre Image Viewer UI, f: Apple Cover Flow UI, and g: Microsoft Vista UI.

In addition, the same seven UI examples were converted to 3D by adding depth information to the image using Adobe Photoshop. The image, including depth information, is sent to the 3DTV, which creates the 3D image, in this case, the 3DUI (see Figure 26).

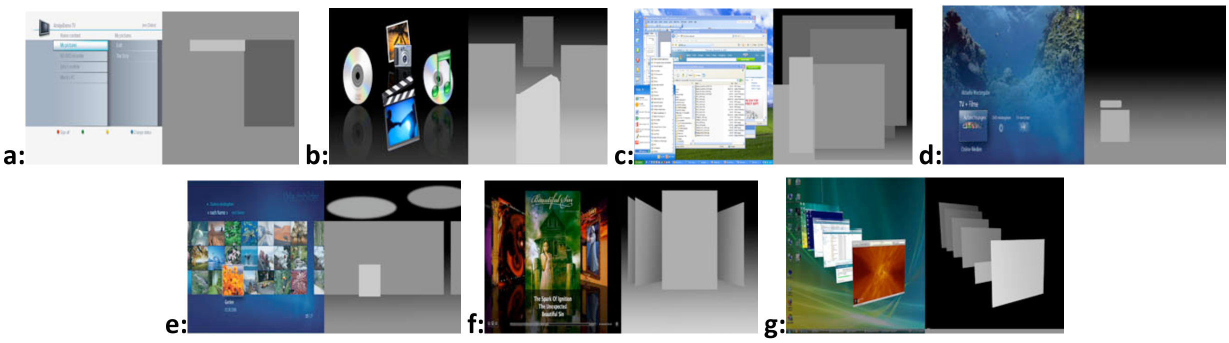


Figure 26: **Pictures of the seven 3DUI examples that were used during the questionnaire of the pilot study (a: Philips Easy Logic TV UI, b: Apple Front Row UI, c: Windows XP UI, d: Vista Media Centre UI, e: Media Centre Image Viewer UI, f: Apple Cover Flow UI, and g: Microsoft Vista UI).**

Procedure

All interviews were conducted from Monday 17th of December 2007 until Thursday 20th of December 2007. Every interview took approximately one hour. The interview started with asking the participants to read and sign a permission form, granting permission to record the interview and to use the results in this project. When signed, the participants were given an introduction to the subject and an explanation of the purpose of this project.

The first part of the interview covered questions about 3D and 3DTV's, 3DUI's, and about gestures as interaction methods (see Appendix A). For some questions, participants had to look at seven screenshots from different UI's (see Figure 25).

Next to this, participants were given a questionnaire and were shown the same seven UI screenshots, but now converted to 3D and displayed on the 3DTV (see Figure 26). All participants were asked to rate the 3DUI examples by filling in the questionnaire, which contained three statements for each example:

- *The 3D effect is very clear!*
- *The 3D effect makes the UI clearer!*
- *The 3DUI invites the use of gestures!*

Participants had to rate using a seven-point Likert scale, ranging from strongly disagree (1) to strongly agree (7), and four (4) being neutral. During these ratings, participants were asked to think aloud so the interviewer could note down any interesting and relevant thoughts. Before the rating started, all the examples were examined, and during rating participants were allowed to scroll back and forth through the examples.

All comments made by the participants were recorded for further analysis. At the same time, the experimenter noted down all answers and comments on plain paper. While noting down the comments, at the same time, the experimenter created a list showing the most frequently given answers per question (see Appendix B). Resembling answers or comments from other participants were easily marked, and new answers were added to the bottom of this list.

3.2 Results

Results from the first part of the pilot study, the interview about 3DTV, 3DUI's, and about Gesture interaction, showed that on the part of 3D, most people see gaming as the next upcoming big application within the world of 3D. Adding to the realism and experience of playing games. One participant even resembled it to the coming of color TV.

"Like the switch from black & white TV to colors, adding more realism to it!"

Nevertheless, there was an overall small consensus about the current lack of quality and other disadvantages, therefore, doubting if 3DTV can live up to its expectations.

"Can it deliver the expectations?"

When looking more specifically at 3DUI's, not much is known about this topic. Overall, most participant foresaw more disadvantages to 3DUI's than they expected an added value. Only one participant mentioned that you could use the extra dimension in the UI to enhance the interface. Furthermore, most participants expected that a 3DUI might be too much, in other words, too complicated to understand, learn, and control.

"With a 3DUI you might create a better conceptual model of the interface's structure!"

Asking what existing UI's would look like when converted to 3D, the participants mainly emphasized the already existing monocular cues. In particular, the monocular cues such as occlusion, blocking the sight of the items which are further away, and relative size, items of the same size as others which look smaller are, hence, further away. Namely, arguing in a way such as *"this item is in front of that, hence it's nearer"*, or *"this item has shades, hence it's in front of the other item"*.

On the part of gesture interaction, simple point and click interaction was the most common mentioned answer. Even so, they emphasized that the arm movements should not be too big or too extensive.

"Not too big, it must be as easy as the remote control is now"

Nevertheless, when participants were actually exposed to the actual 3DUI examples, the different types of gesture interaction they used changed tremendously. For example, now participants tried to grab parts of the interface and tried to push items away. Participants tried to swing items around or flip them over. Participants were clearly more engaged in the interaction and showed more interest in the examples. Some participants even showed a clear change in posture, where in the beginning of the interview they were leaning back in their chairs, now when exposed to the 3DUI examples they where literally sitting on the tip of their chair.

The pilot studies showed that a 3DUI provoked more willingness for gesture interaction compared to what was expected when shown the flat interface. A lack of occlusion in a UI resulted in lower perceived 3D effect, while much occlusion showed high 3D effect. However, this was not always expected up front.

Furthermore, participants often raised the lack of motion during the examples. If the UIs would contain motion, it would most likely enhance their perceived 3D effect, as well as their willingness to interact with the UI by means of gestures.

Besides comments directly linked to the examples, the think aloud method revealed that participants doubted that the 3DTV in a living room setting would be a success because of the current low quality of the 3D screen. Goggles might be a solution to this quality problem. However, this was still not a preferred solution.

"I don't want to put on goggles just to control my TV!"

In addition, they also doubted if gesture control would be the optimum way of interacting with the TV, as they argued that constantly waving your arms would cost too much energy and eventually would be tiresome.

"To constantly wave your arms would be too much of a hassle."

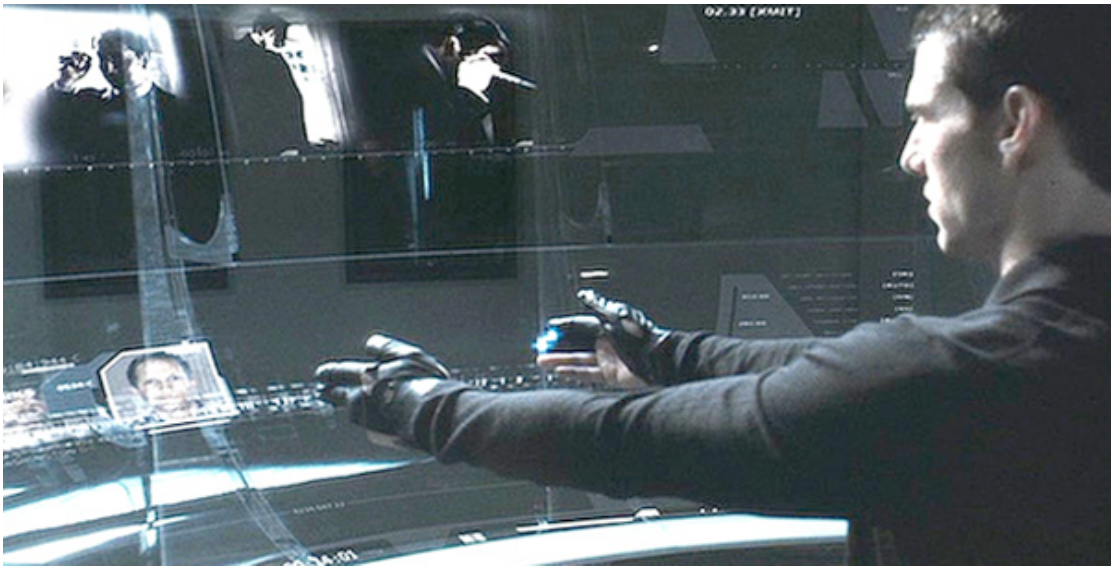


Figure 27: A scene from the Stephen Spielberg's movie Minority Report, showing Tom Cruise making intuitive gestures to interact with vast amounts of data.

Observations from the pilot studies did not show a clear preference for a specific arm movement when participants had to visualize a natural way of interacting. However, most referred to the movie from Stephen Spielberg, Minority Report, and therefore seem anchored by the sort of movements made in that movie (see Figure 27).

"Like in that movie, Minority Report!"

3.3 Discussion

One of the most interesting point from this study might be that the 3DTV and the 3DUI examples provoked more gesture interaction than all participants had imagined up front during the initial interviews. Although this is not supported by hard data, the experimenter clearly observed more interaction and bigger arm movements with almost all participants. Instead of simple point and click gesture interaction, participants now more often wanted to manipulate the 3DUI objects by, grabbing, pushing, pulling, or rotating them. Notably, this also happened with examples which upfront where thought of as not worthy of even converting to 3D. As to the remark that arm movements should not be too big or too extensive, none of such comments were made while interacting with the 3DUI examples.

What provokes these more elaborate and extensive gestures? It could be because of the additional depth cue, as the 3DUI examples were extended with additional binocular depth cue. It could also be that the 3DTV, and hence a 3DUI, has created a spatial metaphor in front of the TV. This spatial metaphor might provoke the participants to create a spatial metaphor for themselves by means of gestures.

Although additional depth cues might provoke more gestures, the lack of even more depth cues might be, in contrast, one of the limitations of this study. For example, the absence of monocular cues such as motion parallax or depth from motion cues. Motion parallax explains that the relative motion of an object gives hints about the relative distance, and depth from motion is determined by dynamically changing object size. In other words, using still, non-moving examples of UI's might be one of the biggest limitations of this study. Without movement, powerful depth cues as mentioned above are excluded from the study.

Take for example the Apple Front Row UI example, initially a strong design which as many participants mentioned to be already very clear, scoring high on “invites the use of gestures”, but scored second lowest on “3D effect”. This might be a perfect example where this limitation seriously affects confidence of the findings, as this Apple UI relies heavily on cues such as motion parallax and relative motion to create its 3D effect. By removing these cues, and although it is converted to 3D and therefore added with binocular cues, the perception of 3D is almost gone.

Apart from the examples being still and non-moving, they also lack any form of interaction and hence feedback. Interaction and in addition visual feedback of that interaction could significantly enhance the participants’ experience.

Therefore, the main experiments should incorporate moving UI element in the examples. Adding more depth cues together with more experience could enhance the 3D effect of the UI and, therefore, provoke more gesture interaction.

4. General Approach

It is important to study how we can keep the interaction with a 3DTV easy and natural, in a way that everybody can understand the system and intuitively interact with it. This chapter states the research questions and hypotheses for this study.

4.1 Research Questions

We have developed some research questions to help ourselves investigate how we can keep the interaction with a 3DTV easy and natural. The following questions represent the focus of the research.

1. *Does a 3DTV enable the user to perceive more depth in a three-dimensional user interface or 3D menu compared to a normal TV, and does it make the interface feel as if it is tangible or graspable?*

   With a 3DUI displayed on a 3DTV, people could have the feeling that the interface or menu is floating in front of the 3DTV as a sort of spatial contexts. They might feel as if the interface is graspable or tangible. It would be interesting to find out if the 3D effect of a 3DTV can be related to that tangible or graspable feeling. Therefore, research is needed to examine to what extent the 3DTV has influence on the perceived 3D effect of a 3DUI.

2. *Does a 3DUI on a 3DTV enhance the willingness to use gesture-based interaction, instead of interacting with more traditional devices such as a remote control?*

   People might feel as if the interface is floating in front of the TV, which creates a spatial metaphor. They might link this spatiality of gestures with the spatial metaphor the 3DUI creates, and as a result they might feel as if gesture interaction is a logic or natural way of interacting with a 3DTV.

4.2 Sub Research Questions

The pilot study showed that motion and occlusion could have a possible interesting influence on people's depth perception and the way they could interact with an interface. Therefore, the difference between binocular cues, motion, and occlusion needs to be incorporated into the main experiment.

1. *Does occlusion within a user interface (which already suggests depth) enhance the perceived 3D effect of the user interface, or the willingness to use gesture interaction?*

2. *Does motion within a user interface (which already suggests depth) enhance the perceived 3D effect of the user interface, or the willingness to use gesture interaction?*

4.3 Hypotheses

Through this research we hope to learn more about how the 3DTV enhances the willingness to use gesture interaction as a natural way of interacting with the TV in the living room. From the research questions presented above we can make the following hypotheses.

1. *It is predicted that all depth cues used in the experiment, binocular depth, motion, and occlusion will have an enlarged effect on the perceived 3D effect of an interface on a 3DTV.*
   Binocular disparity can produce one of the most compelling impressions of depth, and the range and power of occlusion is referred as "striking" (Cutting, 1997). Participant from the pilot study suggested that motion would have significant influence on the 3D effect.
2. *It is predicted that binocular depth on a 3DTV will have an enlarged effect on the willingness to interact with the UI by means of gestures.*
   One of the most interesting points from the pilot study was that the 3DTV and 3DUI examples provoked more gesture interaction than all participants had imagined up front during the initial interviews. Moreover, the experimenter clearly observed more interaction and bigger arm movements with almost all participants.
3. *It is predicted that when an interface consists of binocular depth cues, people will more often choose gestures as an interaction method, prior to more traditional interaction devices, such as the remote control or a keyboard or mouse.*
   Research has shown that one's bare hand is more practical than traditional input devices (Hardenberg & Brard, 2001). Arguably, gesture based interaction could be more practical than interaction with traditional interaction devices.
4. *It is predicted that when an interface consists of binocular depth cues, people will perceive the interface as more graspable compared to an interface without binocular depth cues.*
   Another interesting points from the pilot study was that the 3DTV and 3DUI examples where perceived as more graspable or tangible because participants more often wanted to manipulate the 3DUI objects by, grabbing, pushing, pulling, or rotating them.

5. The Effect of a 3DUI on Interaction Method with a 3DTV

5.1 Introduction

The goal of the main study is to learn how we can interact with a 3DTV in an easy and natural way, such that everybody can understand the system and intuitively interact with it. An experiment will test the binocular depth cues of the 3D-Solution's WOWvx autostereoscopic 3DTV used in a 3DUI, on both the perceived 3D effect of the interface, and on the willingness to use gestures as a natural interaction method. Next to this, the experiment will also test occlusion (monocular cue) and motion as variables of the interface, because participants from the pilot study suggested that these could have substantial influence on the perceived 3D effect or the way people are willing to interact with an interface. In addition, the participants were asked which interaction devices they would prefer to use as an interaction device.

The 3DUI used in the experiment will be based on the Apple Front Row interface because most participants from the pilot study understood the concept of a carrousel as a type of menu. In addition, participants from the pilot study could imagine themselves using gestures in the 2D Apple Front Row example, and the UI scored high on willingness to use gestures in the 3D example.

Because the interface used in the experiment holds a resemblance to the Apple Front Row UI, the experiment will ask participants if they are familiar with Apple products to see if this has an influence. Furthermore, gender, age, level of education and handedness will be recorded.

Furthermore, it could be interesting to observe how participants would interact with the different types of interfaces. Therefore, participants were asked to visually show how they would interact with the 3DUI using gestures, to examine what the most commonly or naturally chosen gesture solutions would be.

Finally, participants were interviewed for their opinion on possible problems of a 3DTV in the living room. For example, problems such as unintended gesturing, who controls the interaction in a family setting, physical effort of gesturing, and the lack of tactile feedback!

5.2 Method

Design

A 2x2x2 within subjects design was employed with motion (with two levels: on/ off), occlusion (with two levels: yes/ no), and depth (with two levels: 2D/ 3D) as independent variable. The dependent variables were subjective assessments of “3D effect” and “ willingness to use gestures”.

Handedness, gender, and level of education were left out because of too little differentiation.

Participants

In total there were 40 participants, all employees of the Philips High Tech Campus, Eindhoven (36 males and 4 females, 38 right and 2 left-handed, 38 technical and 2 none-technical, of which 10 Apple users, ranging from 20-58 *M*: 36.43 *SD*: 10.61). All participants were familiar with the 3DTV in a sense that they had seen it at least once in a demo or product presentation. Participants were briefed about the purpose of the project at the beginning of the interview.

Apparatus and Materials

A 42inch Philips 3D Solution’s 3D WOWvx display (42-3D6W02), connected to a normal personal computer was used to show the different conditions. Participants had to fill in all questionnaires using pen and paper and the whole experiment was recorded using a normal digital HD camcorder.

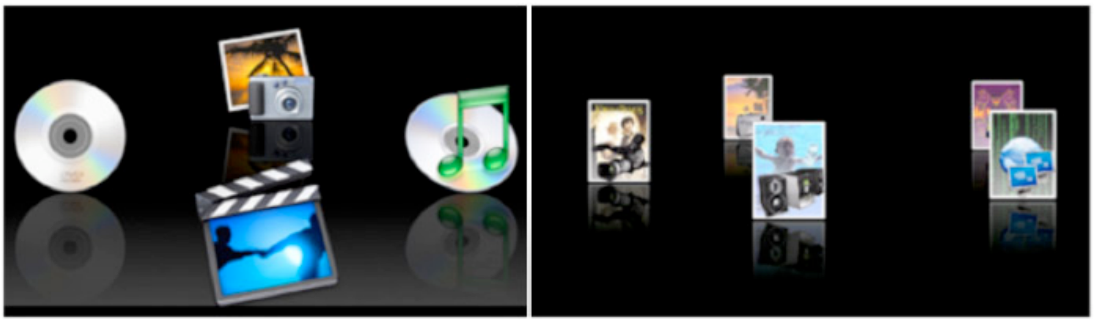


Figure 28: **The Apple Front Row UI (left), and interface design used in this experiment (right).**

The design of the interface was based on the Apple Front Row interface. Figure 28 shows the resemblance and differences between the Apple interface on the left, and the interface used in the experiment on the right. Figures 29-32 show the differences in interface between the eight conditions of the experiment. For more details about the interface used in the experiment see Appendix D.

The first variable, Motion, states that the UI is standing still or that the UI is slowly moving in circles, the Speed and direction of the movement are always fixed. The difference in the condition when motion is turned on/off, is visualized in conditions 5-8 (Figures 31-32) by the white circle illustrating the track the images are moving on.

The second variable, Occlusion, states that some elements of the UI are overlapping each other. Again, they way the element overlap each other is always the same for all conditions. The difference in the condition when occlusion is turned on/off, is visualized in conditions 3-4 (Figure 30), and 7-8 (Figure 32) by some elements overlapped each other in the corners.

The third and last variable, Depth, states that the binocular depth feature of the 3DTV is turned on/off. The difference in the condition when depth is turned on/off, is visualized in conditions 2, 4, 6, and 8 (Figure 29, 30, 31, and 32) by the grayscale figure (depth information) on the right.

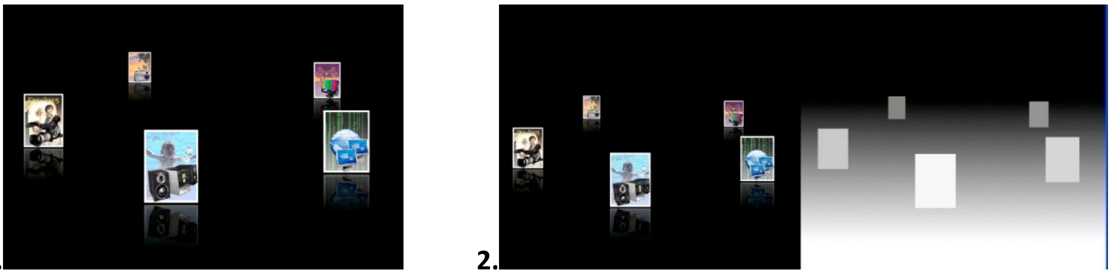
1.  **2.**

Figure 29: **Screenshots of the conditions 1 (the carrousel in 2D, no occlusion & no motion) & condition 2 (the carrousel in 3D, no occlusion & no motion). The grayscale figure on the right side illustrates the depth information that is sent to the 3DTV.**

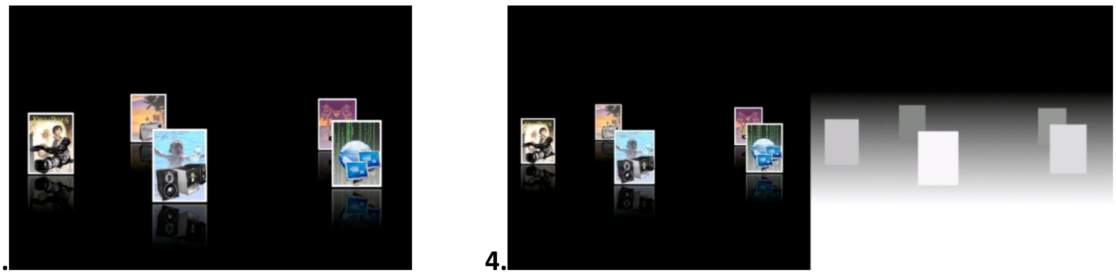
3.  **4.**

Figure 30: **Screenshots of the conditions 3 (the carrousel with occlusion in 2D, no motion) & condition 4 (the carrousel with occlusion in 3D, no motion).**

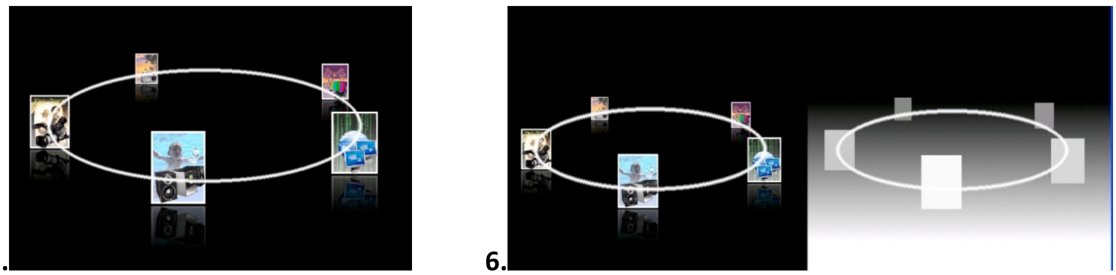
5.  **6.**

Figure 31: **Screenshots of the conditions 5 (the carrousel with motion in 2D, no occlusion) & condition 6 (the carrousel with motion in 3D, no occlusion). The white circle illustrates that the images are moving.**

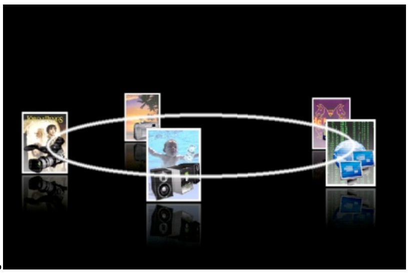
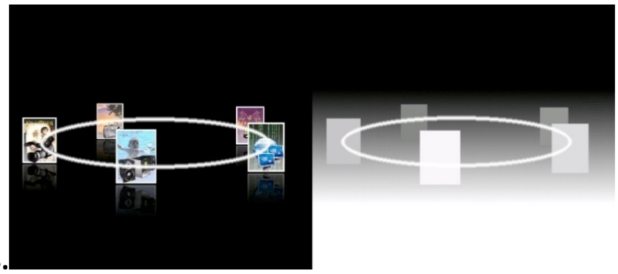
7.  8. 

Figure 32: **Screenshots of the conditions 7 (the carrousel with occlusion & motion in 2D) & condition 8 (the carrousel with occlusion & motion in 3D).**

MEASURES

Participants had to fill in one demographic questionnaire (see Appendix E), covered topics such as age, handedness, gender, and education level.

After each condition (eight times in total), participant had to fill in a questionnaire (see Appendix H) regarding the perceived 3D effect, willingness to use gesture interaction, and their interaction device preference. This questionnaire started by asking to rate the following four statements;

"*The 3D effect is very clear*";
"T*he 3D effect makes the UI clearer*";
"The UI invites the use of gestures";
"I would dislike interacting with this UI using gestures".

Rating each stimulus was done using a seven-point Likert scale ranging from strongly-disagree (1), to strongly-agree (7), with neither disagree nor agree (4) in the middle.

Next to this, participants had to point out what form of interaction they would prefer with the specific condition, choosing from "*not at all*", "*remote control*", "*keyboard and mouse*", "*pointing device*", or "*gestures*".

In addition to the questionnaire, participants were asked to visually show how they would interact with the interface using gestures for each condition. All participants had to visualize how they would select and activate a menu item, and how they would switch between different items. A camera recorded these movements for further analyses.

Afterwards, when all eight conditions were dealt with, the participant was interviewed for their opinion on possible problems of a 3DTV in the living room. The interview contained four problem statements;

"*The problem of unintended gesturing*";

"*The problem of controlling the interaction*";

"The problem of energy cost";

"The problem of no tactile feedback".

The experimenter talked the participants through this final interview and tried to create a lively discussion about the mentioned topics (see Appendix I). During the whole experiment, participants were asked to think aloud and the experimenter noted down any comments or remarks.

PROCEDURE

First, participants were asked to fill in a demographic questionnaire (see Appendix E) and an informed consent form (see Appendix F). Next, participants were briefed about the experiment and were given a frame of reference (see Appendix G), making sure all participants understood the following concepts; quality of the 3DTV, the living room environment, pointing devices and gestures, and a carrousel like interface.

Table 2: **Overview of the eight conditions (motion, occlusion, and depth) used in the experiment.**

Condition	No occlusion		Occlusion	
	2D	3D	2D	3D
No motion	1	2	3	4
Motion	5	6	7	8

Secondly, all participants were exposed to the eight conditions of the experiment (see Table 2). These conditions where counterbalanced over the participants using a "Balanced Latin Square". For every condition, participants were asked to fill in the questionnaire (see Appendix H), and asked to visualize how they would interact with the interface using gestures. Furthermore, after all conditions were shown, participants were interviewed discussing four additional problems (see Appendix I), and the experimenter wrote down any additional comments or remarks.

The time it took to complete one experiment was approximately one hour. Participants from the Philips Consumer Lifestyle ILab were not reimbursed. However, participants from outside the iLab (working somewhere else on the Philips High Tech Campus) were given a Philips Imageo 3 Candlelight set to express our thanks for participating in the experiment.

5.3 Results

The goal of the main study was to learn how we could keep the human 3DTV interaction easy and natural in such a way that everybody can understand the system and intuitively interact with it. This next section will cover the results from the main experiment.

The first part will show the results from the questionnaire, the ratings on perceived 3D effect, willingness to use gesture interaction, and on interaction device preference. The second part shows the results from the observations of what kind of gestures the participants made during the experiment. Thirdly, the results from the answers given during the interview about possible problems in the living room are discussed. Lastly, the outcome of all the comments and reactions that were made during the experiment are summarized.

5.3.1 Part 1: 3D effect, Gesture Willingness, and Interaction Preference

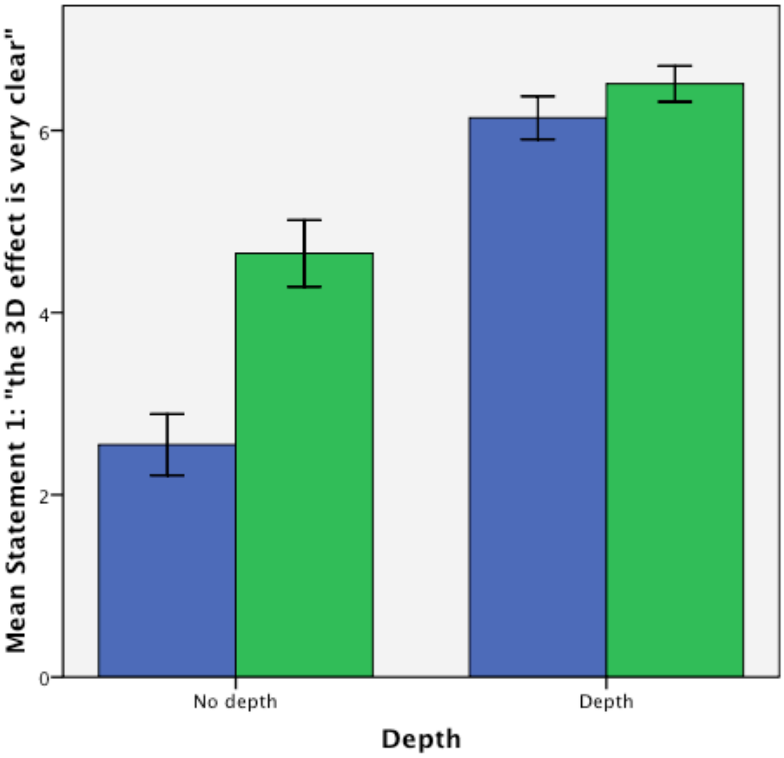
From the questionnaire results, the statements, “*the 3D effect makes the UI clearer*” and *“ I would dislike interacting with this UI using gestures”* were left out.

The first statement was left out because of the resemblance with statement “*the 3D effect is very clear*”. The outcome of both statements was almost the same. The experimenter noted during the rating that participants simply copied their answer from the first to the second statement without really thinking it through. Moreover, some participants did not understand the statement the first time the red it because of the word “clearer”. They had no reference, “clearer than what?”

The second statement was left out because many participants misinterpreted this statement. The experimenter clearly noted that most participants had trouble with the “dislike” expression. The outcome of these results was often counter-intuitive or unbalanced for participants and therefore useless.

Perceived 3D effect

For this part of the study, a repeated measures analysis of variance was performed with motion (on/ off), occlusion (yes/ no), and depth (2D/ 3D) as within-participant factors, and "the 3D effect is very clear" as dependent variable.



	No depth	Depth
No motion	*M* = 2.55 *SD* = 1.52	*M* = 6.140 *SD* = 1.06
Motion	*M* = 4.65 *SD* = 1.65	*M* = 6.51 *SD* = 0.89

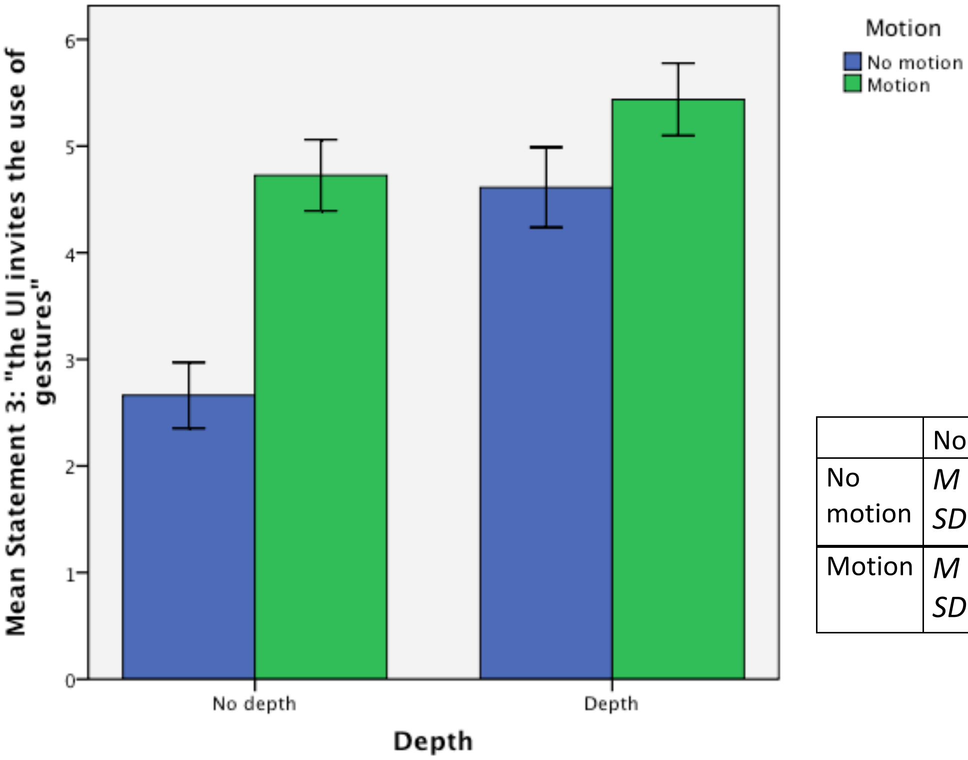
Figure 33: **Interaction effect between motion*depth for *"the 3D effect is very clear"*.**

The test showed a significant main effect for occlusion (*F*(1, 38) = 10.521, p = .002, η = .21), the 3D effect was perceived clearer when the UI consisted of occlusion (*M* = 5.16 *SD* = 2.03) compared to no occlusion (*M* = 4.76 *SD* = 2.03). The test showed a significant main effect for motion (*F*(1, 38) = 145.948, p < .001, η = .79), the 3D effect was perceived clearer when the UI consisted of motion (*M* = 5.58 *SD* = 1.62) compared to no motion (*M* = 4.34 *SD* = 2.22). The test also showed a significant main effect for depth (*F*(1, 38) = 190.369, p < .001, η = .83), the 3D effect was perceived clearer when the UI consisted of depth (*M* = 6.32 *SD* = .99) compared to no depth (*M* = 3.6 *SD* = 1.90).

Furthermore, the test showed a significant interaction effect for motion*depth (*F*(1, 38) = 56.974, p < .001, η = .59). Depth alone is enough to show clear 3D effects, motion only adds to the experience in the no-depth condition (see Figure 33). Other interaction effects were not significant.

WILLINGNESS TO USE GESTURES AS INTERACTION METHOD

A repeated measures analysis of variance was performed with motion (on/ off), occlusion (yes/ no), and depth (2D/ 3D) as within-participant factors, and "the UI invites the use of gestures" as dependent variable.



	No depth	Depth
No motion	*M = 2.66* *SD = 1.39*	*M = 4.61* *SD = 1.69*
Motion	*M = 4.72* *SD = 1.50*	*M = 5.44* *SD = 1.53*

Figure 34: **Interaction effect between motion*depth for *"the UI invites the use of gestures"*.**

The test showed a significant main effect for occlusion ($F(1, 38) = 4.920$, $p = .032$, $\eta = .11$), the UI invited the use of gestures more when the UI consisted of occlusion (*M* = 4.51 *SD* = 1.81) compared to no occlusion (*M* = 4.21 *SD* = 1.87). The test showed a significant main effect for motion ($F(1, 38) = 66.110$, $p < .001$, $\eta = .63$), the UI invited the use of gestures more when the UI consisted of motion (*M* = 5.08 *SD* = 1.55) compared to no motion (*M* = 3.64 *SD* = 1.82). The test also showed a significant main effect for depth ($F(1, 38) = 60.531$, $p < .001$, $\eta = .61$), the UI invited the use of gestures more when the UI consisted of depth (*M* = 5.02 *SD* = 1.66) compared to no depth (*M* = 3.69 *SD* = 1.77).

In addition, the test showed a significant interaction effect for motion*depth ($F(1, 38) = 20.957$, $p < .001$, $\eta = .35$). In this case, motion adds to the willingness to use gestures as interaction method in both the depth and no-depth condition (see Figure 34). Other interaction effects were not significant.

Interaction device preference

For the second part of the questionnaire, for the question "*I would prefer to interact with this interface using: not at all, remote control, keyboard & mouse, pointing device, or gestures*", three crosstabulation chi-square tests were performed for motion (on/ off), occlusion (yes/ no), and depth (2D/ 3D) as independent variables.

Two options were left out because hardly anyone chose these options ("*not at all*" *N* = 4, "*keyboard & mouse*" *N* = 10).

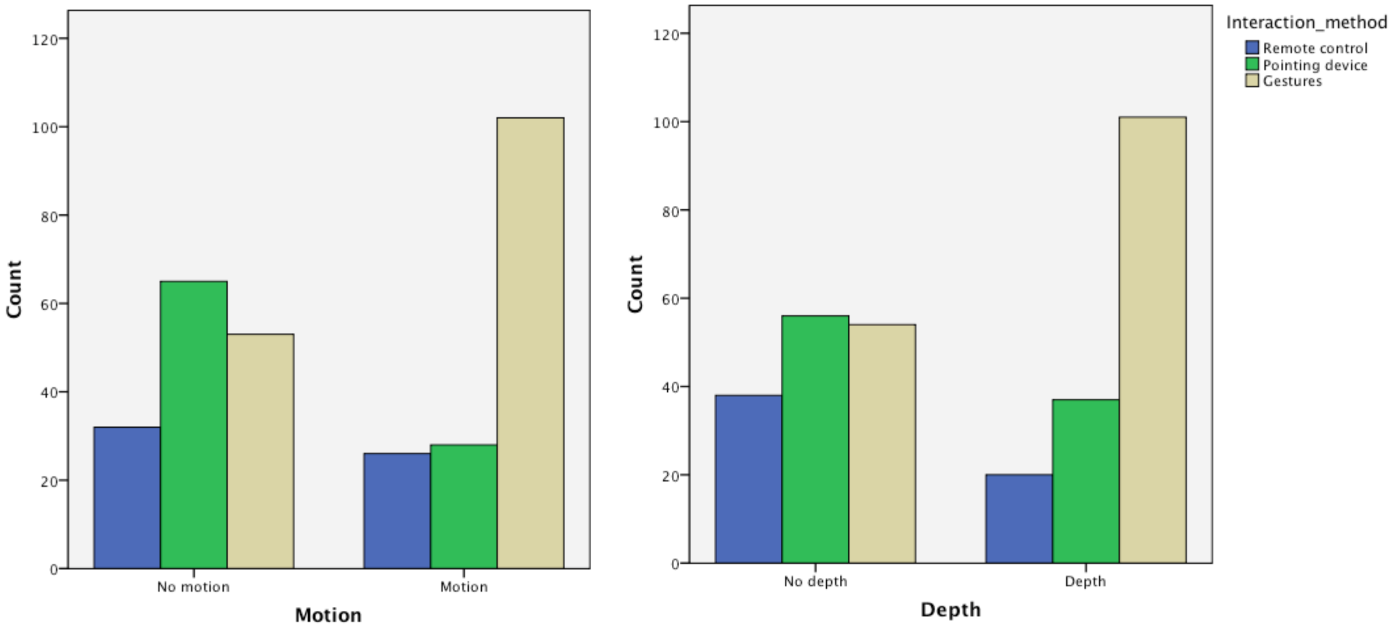


Figure 35: **Bar graph with the count of interaction device preference for motion (left) and depth (right).**

Interaction device preference was significantly affected by motion, X^2 (0.95; 2) = 3.07, p < .001, and by depth, X^2 (0.95; 2) = 2.34, p < .001 (see Figure 35). Gestures were chosen more often when the UI consisted of motion (*N* = 102) compared to no motion (*N* = 53). Likewise, gestures were chosen more often when the UI consisted of depth (*N* = 101) compared to no depth (*N* = 54). Occlusion showed no significant effects on interaction style preference.

5.3.2 Part 2: Observations of the Gestures

After answering the questionnaire, participants were asked to visually show how they would interact with the interface in front of them if they would have to use gestures. The interaction could be divided into two functions, moving the carrousel and selecting an icon. This behavior was recorded during the experiment and analyzed afterwards. One participant was not recorded due to technical reasons. Therefore, only 39 observations are discussed.

Figure 36: **Example of a turning or swinging gesture.**

The results from these observations showed that for moving the carrousel, a flick of the hand, turning or swinging it round, such as seen in Figure 36 was most preferred (N = 23). Grabbing and then turning it came second (N = 7), and some participants showed both these behaviors (N = 5). A small group showed different forms of interaction (N = 4).

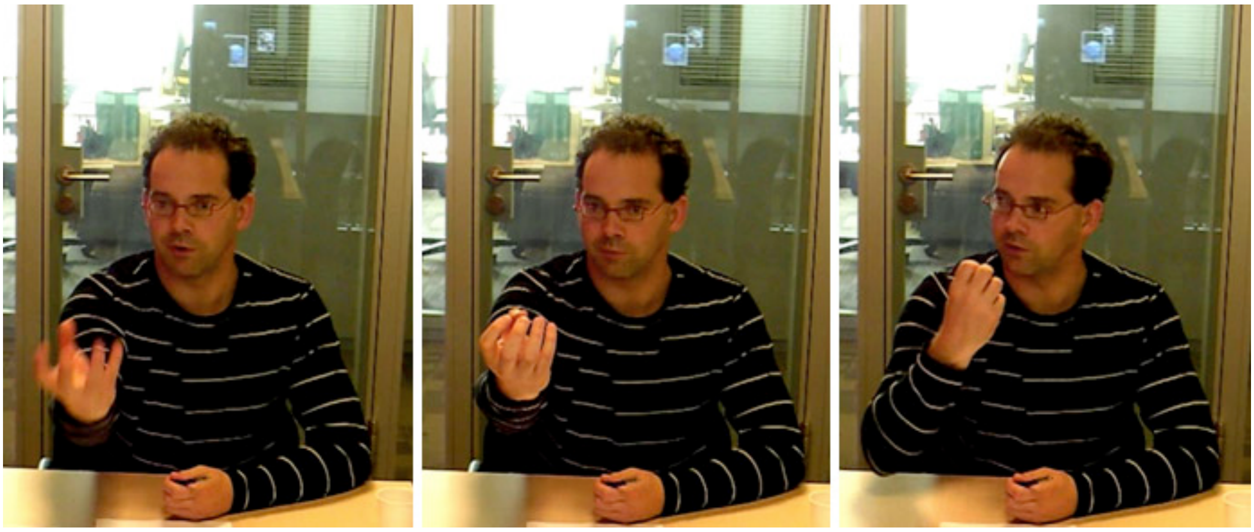


Figure 37: **Example of a grabbing and pulling gesture.**

The results for selecting an icon showed that most (N = 26) participants preferred to press or click the highlighted icon. Grabbing and pulling to the front, such as shown in Figure 37 was chosen second most often (N = 7). Again some participants demonstrated both forms of interaction and only one participant showed different behavior (N = 1).

Clear differences of what type of gestures participants most preferred per experimental condition was not possible to report because after the first condition, participants were biased for the seven following conditions. Making a distinction for only the first conditions was also hard to achieve because participants tried different things and were very exploratory, especially during the first condition. However, when participants first encountered conditions containing no motion and no depth, most made point & click types of interactions and nothing resemble a grabbing type of interaction.

5.3.3 Part 3: Interview about Problems in the Living Room

The next part sums up the most commonly given answers to the four open qualitative questions of the experiment. This part is written to give an overall feel of the participants' reactions. All four questions are discussed separately and the full list of comments can be found in Appendix J.

Problem of unintended gesturing

The first question discusses the problem of unintended gesturing. Gestures or arm movements that are misinterpreted by the TV. For example, when should a menu become active? Should there be a cancel or undo gesture for such mistakes? All participants understood the question and could picture the problem.

The first and most commonly mentioned solution to when such a system should react and listen to gestures was a form of face recognition. The system should only react when the user was looking directly towards the TV. Recognition could be done by either face or head direction. It was argued that this would feel as natural behavior in the same way humans look at each other while interacting.

> *"It is natural to look at somebody when interacting with that person, therefore, it would make sense to apply these same social rules when interacting with a TV, eliminating unintended gestures if you are looking away."*

Secondly, the TV should start listening only after a specific arm movement. Different solutions were given such as, point the finger at the TV, close the hand, a specific code hand signal, a flick of the finger, or even tapping on the knees. All in all, this signs or movements should not resemble a movement that is commonly used in interpersonal communication, but specific for the TV.

A third solution introduced a sort of voice command. An activation command or code word must first be called out before the TV should start listening.

Misinterpretation of a gesture was mainly thought of as unacceptable. However, if this would be the case, some participant suggested different cancel or undo gestures. For example, making a throw away sign with the hands, making a cross with the arms, or shaking with the head.

Problem of controlling the interaction

The second question discusses the problem of who should have control over the interaction. Should the TV listen to everybody or should there be one person in control like with an RC situation? All participants understood the question and could picture the problem.

The overall consent about the problem of who controls the interaction was divided. The majority claimed it to be a social problem. This should not be a concern for the system. One participant talked about TV etiquettes. Discussing that the same problems occurs currently with those who have, or have “stolen” the RC. It is a family matter to sort these forms of hierarchy, not the TV’s problem. Therefore, the TV should react to everybody.

The second group argued that there should only be one person in control at any time, but this control could be passed to somebody else. How this could work, however, remained unclear. For example a child that switched on the TV has gotten control, now the father wants the control back to switch to another channel because he does not like the movie his child is watching. Passing this control from one user to the other, or in this case claiming the control was mentioned as necessary, but unclear how to achieve. Nevertheless, like with a physical RC, this should be easy, fast, and natural.

Apart from the advantage of gesture interaction being a bare hand solution, some participants opted for a physical device that helps the TV to recognize the person in control. Different examples where mentioned such as rings, bracelets, red ball, or even magic wands. All could be easily transferred from one user to the next and the TV could keep track of the person in control. Others suggested having a RC, which is held while making gestures. This RC could then also be used to switch off the gesture mode when watching a movie for example.

Finally, a small group suggested that there should only be one person having control, period. Different suggestions and explications followed, such as de dad, the boss, the man of the house, or even the one who paid for the TV should have ultimate control. There were even ideas to eliminate young children from the chain of control, filtering on hand size or body height. This suggests a sort of user configuration or parental control. Creating different rules for different user.

Problem of physical effort

The third question discusses the amount of energy it costs to constantly interact with the TV through gestures. Is this seen as a problem or might this irritate the user after a while? All participants understood the question and could picture the problem.

The results clearly show that the participants did not see this as a problem because more than twenty participants directly answered this question with a direct no, and only three with a direct yes.

Nevertheless, there are some points that need attention, especially with simple functions such as volume control or channel selection. With these tasks, the interaction should be minimum. Gestures should be recognized when only the forearm or the hands are used. Some even preferred to only use their fingers. Nevertheless, if the elbow must be lifted and full arm movements are needed to interact with the TV, the effort would be too high.

"It should not become a Wii workout!"

However, the main consent was positive. People are used to making gestures in real life. Moreover, when simply watching a TV program, it is not needed to constantly make gestures. Therefore, it might be important to make application content orientated, changing the amount of gestures needed depending on what the user is doing. Maybe integrate a sort of learning system that will accept smaller gestures after a while.

A side effect of gesture interaction, no more RC, was also seen as a positive thing. For example, searching the RC was often mentioned as something that costs a lot of energy and is very irritating. Even getting the RC from the table in front of the couch was envisioned as costing more energy then gesturing. Nevertheless, some participants opt for a dual system, always having a RC at hand.

Finally, social cost was mentioned three times. What would the neighbors think of me if the see me making al kinds of strange arm movements in front of the TV. Apparently, some participants thought that with such a system they would act like or perceived as a fool.

Problem of no tactile feedback

The fourth and final question discusses the problem that gesture interaction lacks the tactile feedback users would normally have if they would interact with a RC. All participants understood the question and could picture the problem.

The results clearly show that the participants did not see the lack of tactile feedback as a serious problem of such a future system. Only three participants found this lack a problem, referring to the tactile feedback currently given by a RC. In addition, one participant emphasized the problem older users could have with such a system; however, it was also argued that they would probably never purchase such a device.

The main consensus states that additive or visual feedback should be enough to forget the missing tactile feedback. With both additive and visual feedback, participants mend to see or hear something as a reaction on their interaction with the system. Additive examples were given in the form of small sounds. Visual feedback was mentioned the most, either in the design of the content, displaying something on the screen. For example stating where and what the status of the gesture is. Or on the TV itself, displaying for example some feedback through a LED indication light. These LEDs should work in the same way the RC indication LED works on current TVs, indicating when the TV is listening to incoming gestures. Additive cues where mentioned the second most as a solution for the missing tactile feedback.

Nevertheless, and this holds for both visual and additive feedback, it is found most important that the feedback reacts directly without any form of delay. If the gesture interaction should come with an unacceptable delay, users would switch back to the RC.

5.3.4 Part 4: Reactions and Comments throught the whole Experiment

This final part sums up the most commonly given reaction and the end and during the experiment (see Appendix K). This part is written to give an overall feel of the participants' reactions.

Overall, half of the participants gave no reactions or extra comments at the end of the experiment, claiming everything to be clear. Others that did have comments mainly emphasized the technical problems they foresaw. Some were afraid that eventually the gestures needed to interact would be too big, and hence would cost too much energy. Furthermore, they could only see a system working when sitting right in front of the TV, not while walking through the living room or to far away from the TV. Some questioned if it would be a need or simply a gadget. Not to mention the costs.

Nevertheless, there were participants interested in such a system, but only if it would work for 100%. And it must work the first time. The 3D binocular cues definitely triggered more gesturing, and it really seemed that people could grab behind the interface.

5.4 Discussion

The main results of the experiment showed that when a UI consists of depth, motion, or occlusion, the 3D effect was perceived clearer compared to a UI without depth, motion, or occlusion. Depth had the strongest effect, followed by motion, and lastly, occlusion having the lowest effect. For willingness to use gestures with the UI, depth, motion, and occlusion had significant influence; all variables invited the use of gesture more compared to a UI without those variables. Again depth had a stronger effect than motion, and occlusion had the smallest effect. The first hypothesis has been accepted as depth, motion, and occlusion have a significant effect on the perceived 3D effect.

The second hypothesis was also accepted because binocular depth had an enlarged effect on the willingness to use gestures as interaction method. In addition, motion and occlusion, though in a lesser extend, had enlarger effects on willingness to use gestures.

The interaction effect of depth and motion; depth had more influence compared to motion regarding the 3D effect of a UI. However, looking at the willingness to use gestures, there was no significant difference between the “no depth & motion” condition and “depth & no motion” condition.

When depth is used in the user interface, people most often chose gestures as interaction method as opposed to the more traditional interaction devices, which is in line with the third hypothesis. However, this was also the case when motion was used in the interface. Occlusion showed no significant difference in the choice of interaction device.

The observations for the type of gestures made, showed that most participants “swung” the carrousel round when they were asked to turn the interface. Nearly one third wanted to grab and then turn the carrousel. When participants were asked to visualize a gesture for selecting an icon, activating by clicking on it was most preferred. With clicking there was a clear resemblance to clicking on the mouse, as most participants also used a form of double click gesture. Again, almost one third of the participants wanted to grab the icon and pull it towards themselves to activate it. Although it is hard to support with data, the experimenter clearly observed graspable and tangible behavior when the UI consisted of binocular depth cues, therefore, carefully accepting the fourth hypothesis.

The main conclusions that can be drawn from the open questions about possible problems of a 3DTV in the living room were that a system should only react to gestures when the user is looking at the TV, or when the user specifically alerts the TV that a gesture is following. Controlling the interaction towards the TV was seen as a social problem and no task for the TV to control. Therefore, the TV should detect gestures from everyone in a living room. However, some want this control to be fixed to one person only, with the possibility to pass the control around. The effort it cost for a person to constantly make gestures was not seen as a problem if the movements where not too big. Finally, the lack of tactile feedback while making gestures was not seen as a big problem because this could easily be covered with visual feedback.

The little reactions participants gave after the experiment concerned the doubt people had about creating a working full proof system, about the technical achievability. Nevertheless, most were enthusiastic and willing to work with such a future system.

The results from this study are in line with the trend towards more natural and intuitive input devices (Subramanian & IJsselsteijn, 2000). Participants from the experiment clearly could envision the 3DTV being the next step in experience in the market of televised entertainment (Fehn, 2006). Nevertheless, most of the participants were also concerned with the negative health effects and picture quality issues. The participants state that these problems need to be solved before the 3DTV can be seen as the logical step following HDTV, as some researchers framed it (Meesters, IJsselsteijn, & Seuntiëns, 2004).

The binocular cues of the autostereoscopic 3DTV clearly enable the user to perceive more depth in the UI. It could be argued that a 3DTV enhances the feeling of presence compared to a normal TV. This supports other research (IJsselsteijn, Ridder, Hamberg, Bouwhuis, & Freeman, 1998; Cutting, 1997) that has shown that adding stereoscopic may enhance the viewer's sense of presence. In addition, and in line with other studies (IJsselsteijn, Oosting, Vogels, Kort, & Loenen, 2006), results from this study show that the cues motion and occlusion also enable the user to perceive more depth in a UI, although in a lesser extent. Moreover, the combination of binocular and monocular cues combined, results in an even bigger 3D effects perceived in the interface.

Regarding gesture interaction, the results are in line with previous research (Hardenberg & Brard, 2001) stating that bare hand interaction is more practical than traditional input devices. It shows that most participants choose gesture-based interaction over the remote control to interact with a 3DUI on a 3DTV. Furthermore, it was expected that deictic gestures could be useful for human 3DTV interaction because they are useful for human computer interaction (Bolt, 1980; Lenman, 2002). Our study indeed shows that gesture based interaction and deictic gestures are useful for human 3DTV interaction, and therefore, could also be useful in the living room.

There are several factors that need to be considered in evaluating the findings of the present research. First, although most conditions consisting of depth showed positive significant results, outcomes might have been different if the experiment was not limited to one specific type of 3DTV. Comments from the think aloud method revealed a negative tone regarding the quality and resolution of the 3DTV. In addition, the limited maximum depth and few viewing angles could have also had influence on the ratings. Further study should try to minimize the difference between the normal TV and the 3DTV, possibly by using a higher quality 3DTV or downgrade the image quality of the normal TV.

Furthermore, the participants used in this experiment where all employees working at the Philips High Tech Campus and most where men. Almost all qualified themselves as being technical and finished some form of higher education. Hence, participants were not representative of the general population; a further study should include more woman, non-technical, and lower educated participants. Nevertheless, even with this group of technical participants, many had problems envisioning how to interact with the UI using gestures because for most this was a fairly new concept or they had never thought about actually using it.

Finally, the third problem with this study was that participants had to envision the gesture interaction. This means that the experimenter might have influenced or anchored the participants because some explanation and examples were necessary to introduce gestures interaction. A future study should try to include a working prototype of gesture interaction so that participants can explore the differences between the choices without the possibility of being influenced by the experimenter. However, less technical people would have a much larger task envisioning gesture-based interaction.

6. Conclusion

The main results of the experiment show that the binocular depth cues of an autostereoscopic 3DTV can enhance the 3D effect of a 3DUI, and also enhance the willingness to use gesture interaction with the interface. In the 3DUI, depth cues motion and occlusion also had influence on the 3D effect, but to a lesser extent. When the interface consists of binocular depth cues, participants preferred gesture interaction above traditional interaction devices.

The present study demonstrates that the key feature of an autostereoscopic lenticular 3DTV, binocular depth cues, have a large influence on the perceived 3D effect of a UI. As a consequence, binocular depth cues have a large influence on the willingness to interact with a UI using gestures. It could appear that such a 3DUI creates a spatial metaphor in front of the TV, enhancing the willingness to use gestures instead of traditional interaction devices. Many participants described this as the UI being more "graspable".

The acceptance and familiarity of gesture interaction combined with such interfaces, however, is currently still a problem. Many participants needed additional explanation from the experimenters to envision themselves interacting by means of gestures. Even so, this might all change rapidly. Already, most participants familiar with Apple products showed a better understanding of the type of interface and the envisioned way of gesture interaction compared to none Apple users.

Former research has concluded that gesture based research has been done in theory, yet little has been applied (Karam & Schraefel, 2005), nevertheless, looking at the near future of human computer interaction, we expect large changes in the upcoming years.

Figure 38: **Examples of multiple US broadcasters reporting on the presidential elections using massive HD multi touch display solution such as the Perspective Pixel's Media Wall (Engadget, 2008).**

Apple´s enormous iPhone hype introduced massive amounts of people to multi finger intuitive gesture interaction. Microsoft's newly announced Windows 7 Operating System; released in 2009 will be fully touch and gesture based. All these developments will enhance the acceptance and familiarity of natural ways of gesture interaction.

Furthermore, the media will also help with the acceptance of intuitive and natural interaction techniques. For example (see Figure 38), the US 2008 presidential elections were covered by most of the major US broadcasters using massive ZUI touch screen displays. Movies such as James Bond, Quantum of Solace, also show intuitive gesture interaction on big touch screen tabletop displays, such as displayed on Figure 39.

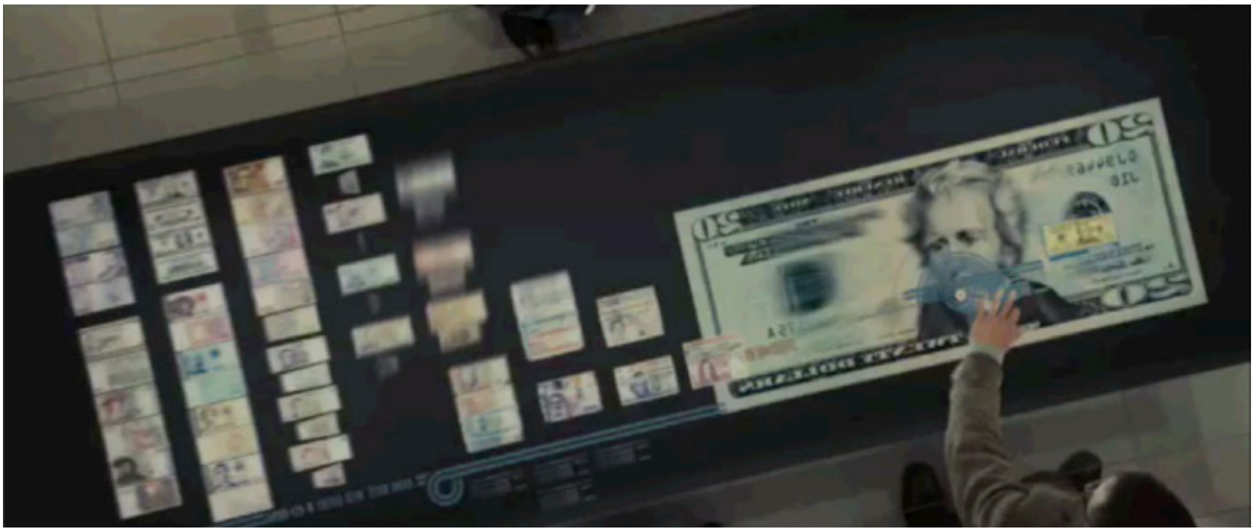


Figure 39: **A Minority Report/iPhone/Microsoft Surface type large-scale interface featuring multi-touch, gesture- control and magical-connection-with-every-database-and-communication-system-on-earth (Cnet News, 2008).**

Although most of these gestures will still rely on physically touching a screen, it is an initial step away from the traditional desktop interface interaction and opens the door for more natural and intuitive ways of interacting. A final step is to move away from the touch screen and into the open space, for example interacting with the TV from the couch in the living room. Game console manufacturers will continue on the success of the Nintendo Wii's interface, and probably also venture into full body interaction to create more natural interactions.

One of the best examples of intuitive gesture interaction, which was mentioned by most participants in the experiments, is from Steven Spielberg´s Hollywood movie Minority Report. Companies such as Oblong (Oblong, 2008) are working on real world implementations of such systems, and we might see more of such examples in the future.

Nevertheless, this study shows that binocular cues enhance the willingness to use gesture interaction with a user interface. Therefore, intuitive gesture interaction from a distance with the TV in the living room, might just work when the interface is displayed on a 3DTV.

The 3DTV's binocular depth cues might just be the thing people need to cross the line and start interacting with the TV from a distance using intuitive 3D gestures.

References

Agarawala, A., & Balakrishnan, R. (2006). Keepin' It Real: Pushing the Desktop Metaphor with Physics, Piles and the Pen. *Proceeding of CHI'06*, (pp. 1283-1292).

Alternative Film Guide. (2008, March 12th). *US Movie Theaters Going 3D*. (A. Soares, Editor) Retrieved August 16th, 2008, from Alternative Film Guide: http://www.altfg.com/blog/film-movies/us-movie-theaters-going-3d/

Apple Inc. (2008). *iPhone*. Retrieved August 16th, 2008, from Apple.com: http://www.apple.com/iphone/

BBC News. (2008, March 10th). *Broadcasting first with 3D rugby*. Retrieved August 16th, 2008, from BBC News: http://news.bbc.co.uk/2/hi/technology/7286852.stm

Benko, H., Wilson, A., & Baudish, P. (2006). Precise Selection Techniques for Multi-Touch Screens. *Proceedings of CHI ’06, Human Factors in Computing Systems*, (pp. 1263-1272).

Benton, S. (2001). Selected Papers on Three-Dimensional Displays. *SPIE Optical Engineering Press* .

Blackwell, A., Fitzmaurice, G., Holmquist, L., Ishii, H., & Ullmer, B. (2007). Tangible user interfaces in context and theory. *Proceedings of CHI ’07* , 2817-2820.

Bliss, J., Tidwell, P., & Guest, M. (1997). The Effectiveness of Virtual Reality for Administering Spatial Navigation Training to Firefighters. *Presence Teleoperators & Virtual Environments , 6* (1), 73-86.

Blundell, B., & Schwarz, A. (2005). *Creative 3-D Display and Interaction Interfaces: A Trans-Disciplinary Approach.* Wiley, John & Sons, Incorporated.

Bolt, R. (1980). “Put-That-There”: Voice and Gesture at the Graphics Interface. *SIG-. GRAGH ’80 Proceedings* , 262-270.

Bookshire Conner, D., Snibbe, S., Herndon, K., Robbins, D., Zeleznik, R., & Dam, A. v. (1992). Threedimensional widgets. *ACM Symposium on Interactive 3D Graphics* , 183-188.

Bowman, D., Kruijff, E., LaViola, J., & Poupyrev, I. (2004). *3D User Interfaces: Theory and Practice. .* Addison Wesley.

Bryson, S. (1996). Virtual Reality in Scientific Visualization. *Communications of the ACM , 39* (5), 62-71.

Butterworth, J., Davidson, A., Hench, S., & Olano, M. (1992). 3DM: A Three-Dimensional Modeler Using a Head Mounted Display. *Proceedings of the 1992 ACM Symposium on Interactive 3D Graphics (I3D '92).* ACM Press.

Cadoz. (1994). *“Les réalités virtuelles,” Dominos, Flammarion. .*

Canesta. (n.d.). *TOUCH-FREE GESTURE CONTROL WILL NET WINNING PRODUCTS, INCREASED MARKET SHARE & REVENUES FOR PARTICIPATING TV & PC MAKERS* . Retrieved March 18th, 2009, from Canesta.com: http://www.canesta.com/corporate/press-releases/031109

Carlin, A., Hoffman, H., & Weghorst, S. (1997). Virtual Reality and Tactile Augmentation in the Treatment of Spider Phobia: A Case Report. *Behavior Research and Therapy , 35* (2), 153-159.

Cnet Australia . (2008, Januari 10th). *TV comes to life with gesture technology*. (B. Winterford, Editor) Retrieved November 16th, 2008, from Cnet Australia: http://www.cnet.com.au/tvs/0,239035250,339285008,00.htm

Cnet news. (2008, November 17th). *The Gadget-less Bond*. (A. Richardson, Editor) Retrieved November 18th, 2008, from Cnet news : http://news.cnet.com/matter-antimatter/?authorId=9759474

Cruz-Neira, C., & Lutz, R. (1999). Using Immersive Virtual Environments for Certification. *IEEe Software , 16* (4), 26-30.

Cutting, J. (1997). How the eye measures reality and virtual reality. *Behavior Research Methods, Instruments & Computers , 29*, 27-36.

Dodgson, N. (1996). Analysis of the viewing zone of the Cambridge autostereoscopic display. *Applied Optics , 35*, 1705-1710.

Dodgson, N. (Ed.). (2005). Autostereoscopic 3D Displays. *Proceedings of the IEEE Computer , 38* (8), 31-36.

Duh, H., Parker, D., Phillips, J., & Furness, T. (2004). 'Conflicting' motion cues to the visual and vestibular self-motion systems around 0.06 Hz evoke simulator sickness. In *Human Factors* (Vol. 46, pp. 142-153).

Engadget HD. (2008, April 9th). *Ask Engadget HD: When will wireless HD / HDMI take off?* (D. Murph, Editor) Retrieved August 16th, 2008, from Engadget HD: http://www.engadgethd.com/2008/04/09/ask-engadget-hd-when-will-wireless-hd-hdmi-take-off/

Engadget HD. (2008, July 31st). *DISH to have 150 HD channels this year and 1080p VOD in August*. (B. Drawbaugh, Editor) Retrieved August 16th, 2008, from Engadget HD: http://www.engadgethd.com/2008/07/31/dish-to-have-150-hd-channels-this-year-and-1080p-vod-in-august/

Engadget. (2008, January 6th). *Miracube G240S ushers in "new generation" of 3D displays*. (B. Dybwad, Editor) Retrieved August 16th, 2008, from Engadget: http://www.engadget.com/2008/01/06/miracube-g240s-ushers-in-new-generation-of-3d-displays/

Engadget. (2008, November 14th). *Oblong's g-speak: the 'Minority Report' OS brought to life*. (J. Topolsky, Editor) Retrieved November 16th, 2008, from Engadget: http://www.engadget.com/2008/11/14/oblongs-g-speak-the-minority-report-os-brought-to-life/

Engadget. (2008, August 29th). *Panasonic's 150-inch Life Wall TV going production in 2009*. (T. Ricker, Editor) Retrieved August 29th, 2008, from Engadget: http://www.engadget.com/2008/08/29/panasonics-150-inch-life-wall-tv-going-production-in-2009/

Engadget. (2008, August 28th). *Philips' 8mm thin 32-inch LCD turns the screws on Sony*. (T. Ricker, Editor) Retrieved August 29th, 2008, from Engadget: [6] http://www.engadget.com/2008/08/28/philips-8mm-thin-lcd-turns-the-screws-on-sony/

Engadget. (2009, May 1st). *Xbox 360 to get motion-sensing add-on with full body game control?* (P. Miller, Editor) Retrieved May 6th, 2009, from Engadget: http://www.engadget.com/2009/05/01/xbox-360-to-get-motion-sensing-add-on-with-full-body-game-contro/

EngadgetHD. (2008, November 4th). *Election '08 coming home in HD like never before*. (R. Lawier, Editor) Retrieved November 4th, 2008, from EngadgetHD: http://www.engadgethd.com/2008/11/04/election-08-coming-home-in-hd-like-never-before/

Fehn, C. (2004). 3D-TV using Depth-Image-Based Rendering (DIBR). *Proceedings of Picture Coding Symposium* .

Fehn, C. (2006). Interactive 3-DTV-concepts and Key Technologies. *Proceedings of the IEEE , 94* (3), 524-538.

Figueroa, P., Dachselt, R., & Lindt, I. (2006). A Conceptual Model and Specification Language for Mixed Reality Interface Components. *Proceedings of the Workshop 'Specification of Mixed Reality User Interfaces: Approaches, Languages, Standardization.* Alexandria (Virginia, USA): IEEE Computer Society.

Fitzmaurice, G., Ishii, H., & Buxton, W. (1995). Bricks: Laying the foundations for graspable user interfaces. *Proceedings of the ACMSIGCHI Conference on Human Factors in Computing Systems (CHI'95)* , 442–449.

Forsberg, A., Kirby, M., Laidlaw, D., Karniadakis, G., Dam, A. v., & Elion, J. (2000). Immersive Virtual Reality for Visualizing Flow though an Artery. *Proceedings of IEEE Visualization 2000.*

GameCyte. (2009, Jan 7th). *Is Your Gaming Future in Stereo 3D? — Part Five*. (S. Hollister, Editor) Retrieved May 6th, 2009, from GameCyte: http://www.gamecyte.com/is-your-gaming-future-in-stereo-3d-part-five

Gulfnews. (2008, April 27th). *Coming home to HDTV*. (A. Anastasiou, Editor) Retrieved August 16th, 2008, from Gulfnews: http://archive.gulfnews.com/gnfocus/av_system_sintegration_2008/sub_story/10208686.html

Häkkinen, J., Liinasuo, M., Takatalo, J., & Nyman, G. (2005). Visual comfort with stereoscopic mobile gaming. *Proceedings of the IS&T/SPIE's International Symposium on Electronic Imaging , 6055*.

Han, J. (2006, August). *Unveiling the genius of multi-touch interface design*. Retrieved August 16th, 2008, from TED ideas worth spreading - Talks: http://www.ted.com/index.php/talks/jeff_han_demos_his_breakthrough_touchscreen.html

Hardenberg, C., & Brard, F. (2001). Bare-hand human-computer interaction. In *Proceedings of the AMC workshop on Perceptive user interfaces.* New York: ACM Press.

HDTV News. (2008, May 14th). *Samsung to demo 240-Hz LCD TV panel*. (J. Harris, Editor) Retrieved August 16th, 2008, from HDTV News: http://www.hdtv-news.co.uk/2008/05/14/samsung-to-demo-240-hz-lcd-tv-panel/

Hodges, L., Kooper, R., Meyer, T., Rothbaum, B., Opdyke, D., de Graaff, J., et al. (1995). Virtual Environments for Treating the Fear of Heights. *IEEE Computer , 28* (7), 27-34.

Hodges, L., Rothbaum, B., Alarcon, R., Ready, D., Shahar, F., Graap, K., et al. (1995). Virtual Vietnam: A Virtual Environment for the Treatment of Chronic Post-traumatic Stress Disorder. *CyberPsychology and Behavior , 2* (1), 7-14.

Hopf, K. (2006). Novel Autostereoscopic Single-User Displays with User Interaction. *Proceedings of the SPIE , 6392*.

HP. (2008). *HP Smarttouch PC, touch the future now. Retrieved*. Retrieved November 13th, 2008, from Engadget: http://www.hp.com/united-states/campaigns/touchsmart/#/Main/

Hur, N., Ahn, C., & Ahn, C. (2002). Experimental service of 3DTV broadcasting relay in Korea. *Proc. SPIE*, *4864.*

IJsselsteijn, W., Oosting, W., Vogels, I., Kort, Y. d., & Loenen, E. (2006). Looking At or Looking Out: Exploring Monocular Cues to Create a See-Through Experience with a Virtual Window. *Proceedings of Presence '06* , 87-95.

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

IJsselsteijn, W., Seuntiens, P., & Meesters, L. (2005). Human factors of 3D displays. In P. K. O. Schreer, *3D Videocommucication - Algorithms, concepts and real-time systems in human-centred communication* (pp. 219-234). New Jersey: John Wiley & Sons.

Ishii, H., & Ullmer, B. (1997). Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms. *Proceedings of CHI '97* , 234-241.

ITEA2. (2008). *About ITEA 2*. Retrieved April 2008, from ITEA 2, Infromation Technology for European Advancement: http://www.itea2.org/

Jacob, R., Leggett, J., Myers, B., & Pausch, R. (1993). Interaction styles and input/output devices. *12* (2), 69-79.

Joseph, J., & LaViola, J. (2008). Bringing VR and Spatial 3D Interaction to the Masses through Video Games. *IEEE Comput Graph Appl. , 28* (5), 10-15.

Karam, M., & Schraefel, M. (2005). *A Taxonomy of Gestures in Human Computer Interactions.* Electronics and Computer Science, University of Southampton. (Submitted).

Kasik, D., Troy, J., Amorosi, S., Murray, M., & Swamy, S. (2002). Evaluating Graphics Displays for Complex 3D Models. *IEEE Computer Graphics and Applications , 22* (3), 56-64.

Kennedy, R., Lanham, D., Drexler, J., Massey, C., & Lilienthal, M. (1997). A comparison of cybersickness incidences, symptom profiles, measurement techniques, and suggestions for further research. In *Presence: Teleoperators and Virtual Environments* (Vol. 6, pp. 638-644).

Kirsh, D. (1995). Complementary strategies: why we use our hands when we think. *Proceedings of 7th Annual Conference of the Cognitive Science Society* , 212-217.

Kirsh, D., & Maglio, P. (1994). On Distinguishing Epistemic from Pragmatic Action. *Cognitive Science 18 , 4*, 513-549.

Kitamura, Y., Konishi, T., Masaki, T., & Kishino, F. (2001). IllusionHole: a stereoscopic display for multiple observers. *Proceedings of the SPIE*, *380*, p. 8.

Kolasinski, E. (1995). *Simulator sickness in virtual environments (ARI Technical Report 1027).* U.S. Army Research Institute for the Behavioral and Social Sciences.

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

Lambooij, M., IJsselsteijn, W., & Heynderickx, I. (2007). Stereoscopic displays and visual comfort: a review. In *Stereoscopic displays and virtual reality systems XIV.*

Lenman, S. (2002). Computer Vision Based Hand Gestures Interfaces for Human-Computer Interaction. *Technical report TRITANA -D0209 , 172*.

Liang, J., & Green, M. (1994). JDCAD: A Highly Interactive 3D Modeling System. *Computers and Graphics , 18* (4), 499-506.

Mark, W. (1999). *Post Rendering 3D Image Warping: Visibility, Reconshuction and Performance for Depth Image Warping.* PhD Thesis, University of North Carolina, Chapel Hill.

McAllister, D. (Ed.). (1993). *Stereo Computer Graphics and Other True 3D Technologies.* Princeton Univ. Press.

McMillan, L. *An Image-Based Approach to Three-Dimensional Computer Graphics.* PhD Thesis, University of North Carolina, Chapel Hill, 1997.

McNeill, D. (1992). *Hand and mind: what gestures reveal about thought.* Chicago, USA: University of chicago press.

Meesters, L., IJsselsteijn, W., & Seuntiëns, P. (2004). A survey of perceptual evaluations and requirements of three-dimensional TV. *IEEE Transactions on Circuits and Systems in Video Technology , 14*, 381-391.

Microsoft Corporation. (2008). *Experience Surface*. Retrieved August 16th, 2008, from Microsoft Surface: http://www.microsoft.com/surface/index.html

Multichannel News. (2007, July 8th). *NBC to Shoot Record 3,600 Hours in Beijing for Olympics*. (S. Donohue, Editor) Retrieved August 16th, 2008, from Multichannel News: http://www.multichannel.com/article/CA6466377.html

Multichannel News. (2008, August 9th). *Q&A: AT&T's York Scopes Out HD Plans*. (T. Spangler, Editor) Retrieved August 16th, 2008, from Multichannel News: http://www.multichannel.com/article/CA6586004.html?industryid=47194

Nespoulous, J., Perron, P., & Lecours, A. (1986). *The Biological Foundations of Gestures: Motor and Semiotic Aspects.* Hillsdale, MJ: Lawrence Erlbaum Associates.

Okamura, S. (1994). *History of Electron Tubes.* IOS Press.

Okoshi, T. (1976). *Three Dimensional Imaging Techniques.* New York: Academic Press.

Patten, J., Ishii, H., Hines, J., & Pangaro, G. (2001). Sensetable: A Wireless Object Tracking Platform for Tangible User Interfaces. *Proceedings of CHI '01.*

Pausch, R., Snoddy, J., Taylor, R., Watson, S., & Haseltine, E. (1996). Disney's Aladdin: First Steps toward Storytelling in Virtual Reality. *Proceedings of SIGGRAPH '96.* ACM Press.

Philips. (2007, October). *42" Philips 3DWOW 3D Display - WOWvx Technology (Model: 42-3D6W02).* Retrieved August 16th, 2008, from Philips: http://www.business-sites.philips.com/shared/assets/global/Downloadablefile/Philips-3D-Solutions---42inch-leaflet-(version-October-07)(1)-14028.pdf

Philips. (2008). *About.* Retrieved August 16th, 2008, from Home, 3D Solutions: http://www.business-sites.philips.com/3dsolutions/about/index.page

Philips. (2008). *Businesses.* Retrieved August 16th, 2008, from Home, Company and Press, Company Profile: http://www.philips.com/about/company/businesses/index.page

Philips. (2008). *Consumer products, Televisions, TVs*. (Philips, Editor) Retrieved August 16th, 2008, from Philips sense and simplicity: http://www.consumer.philips.com/consumer/en/gb/consumer/cc/_categoryid_TV_CA_GB_CONSUMER/

Philips News Center. (2008, June 24th). *First commercial rollout Ambient Experience with citizenM hotels pioneers new direction for hospitality industry*. Retrieved August 16th, 2008, from About Philips : http://www.newscenter.philips.com/about/news/press/20080624_citizen_m.page

Philips News Center. (2008, Juni 10th). *NH Hoteles and Philips partner for two year study into consumer well-being*. Retrieved August 16th, 2008, from About Philips: http://www.newscenter.philips.com/about/news/press/20080610_nh_hotel.page

Planar Systems Inc. (2009). *StereoMirror*. Retrieved Februari 12th, 2009, from Planar: When image experience matters: http://www.planar3d.com/

ReadersDigest.com. (n.d.). *Become a One-Television Household*. (S. Zimmerman, Editor) Retrieved March 18th, 2009, from ReadersDigest.com: http://www.rd.com/your-america-inspiring-people-and-stories/for-america-become-a-one-television-household/article29482.html

Rime, B., & Schiaratura, L. (1991). *"Gesture and speech". Fundamentals of Nonverbal Behaviour.* New York: Press Syndacate of the University of Cambridge.

Sand, R. (1992). 3-DTV-a review of recent and current developments. *IEE Colloquium on Stereoscopic Television* , 1-4.

Schönhage, B., & Eliëns, A. (1999). Dynamic and mobile vrml gadgets. *VRML '99: Proceedings of the fourth symposium on Virtual reality modeling language* (pp. 47–52). New York, NY, USA: ACM Press.

SenterNovem. (2008). *About us*. Retrieved 4 16, 2008, from Your Dutch partner for innovation and sustainable development: http://www.senternovem.nl/english/index.asp

Subramanian, S., & IJsselsteijn, W. (2000). Survey and Classification of Spatial Object Manipulation Techniques. *Proceedings of OZCHI 2000, Interfacing Reality in the New Millennium*, (pp. 330-337).

TechRadar.com. (2008, Feb 2nd). *Gaming News*. (A. Hartley, Editor) Retrieved May 6th, 2009, from Stereoscopic 3D games for PS3 and Xbox 360: Blitz Games unveils 3D HD tech at 3D Entertainment Summit: http://www.techradar.com/news/gaming/stereoscopic-3d-games-for-ps3-and-xbox-360-490089

Telegraph.co.uk. (2008, October 28th). *Microsoft shows off Windows 7 operating system* . (C. Elsworth, Editor) Retrieved November 13th, 2008, from Telegraph.co.uk: http://www.telegraph.co.uk/scienceandtechnology/3358946/Microsoft-shows-off-Windows-7-operating-system.html

Tweakers. (2008, November 16th). *Besturingssysteem uit Minority Report gedemonstreerd*. (d. W. Moor, Editor) Retrieved November 16th, 2008, from Tweakers : http://tweakers.net/nieuws/56790/besturingssysteem-uit-minority-report-gedemonstreerd.html

U23D. (2008). *About the Film*. Retrieved August 16th, 2008, from U23Dmovie: http://www.u23dmovie.com/

Wikipedia. (2008, July). *Anaglyph image*. Retrieved August 16th, 2008, from Wikipedia, the free encyclopedia: http://en.wikipedia.org/wiki/Anaglyph_image

Wilson, A. (2005). PlayAnywhere: A Compact Interactive Tabletop Projection-Vision System. *Symposium on User Interface Software and Technology (UIST).*

Wilson, A. (2004). TouchLight: An Imaging Touch Screen and Display for Gesture-Based Interaction. *International Conference on Multimodal Interfaces.*

Appendixes

A. Pilot Study: Examples and Questions

This Appendix contains the user interface examples used during the interview and questionnaire of the pilot study. Furthermore, it will enclose the questions used during the interview and questionnaire.

UI Examples

The following seven examples were used during the interview part of the pilot study.

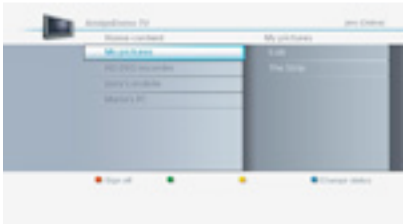


Figure 40: **Philips Easy Logic TV UI**

Figure 41: **Apple Front Row UI**

Figure 42: **Windows XP UI**

Figure 43: **Vista Media Centre UI**

Figure 44: **Media Centre Image Viewer UI**

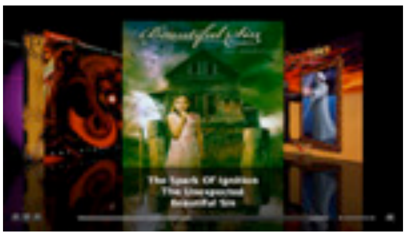


Figure 45: **Apple Cover Flow UI**

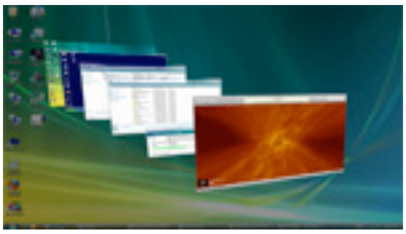


Figure 46: **Microsoft Vista UI**

3DUI EXAMPLES

The following seven examples were used during the questionnaire part of the pilot study.

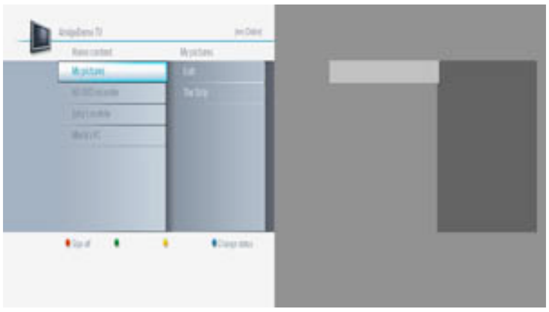


Figure 47: **Philips Easy Logic TV 3DUI**

Figure 48: **Apple Front Row 3DUI**

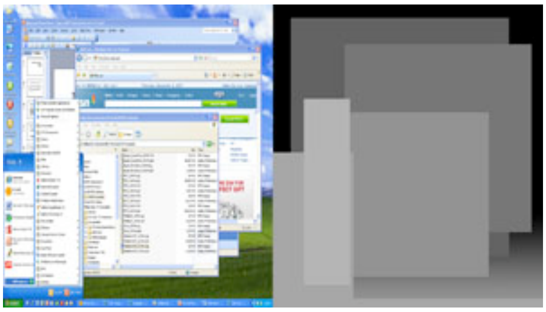


Figure 49: **Windows XP 3DUI**

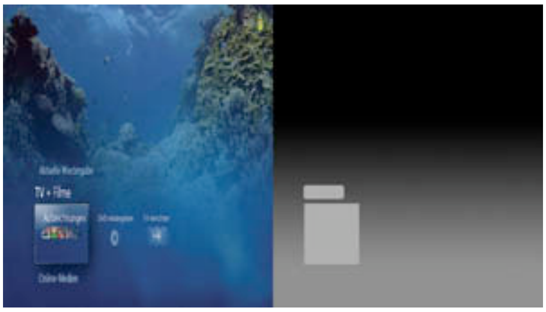


Figure 50: **Vista Media Centre 3DUI**

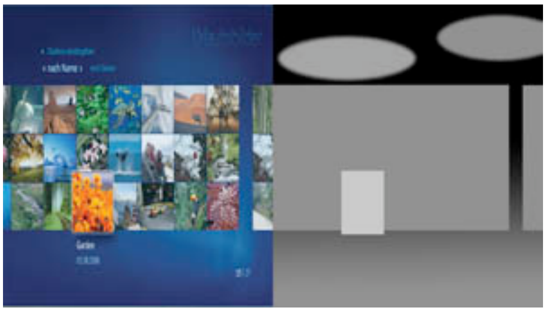


Figure 51: **Media Centre Image Viewer 3DUI**

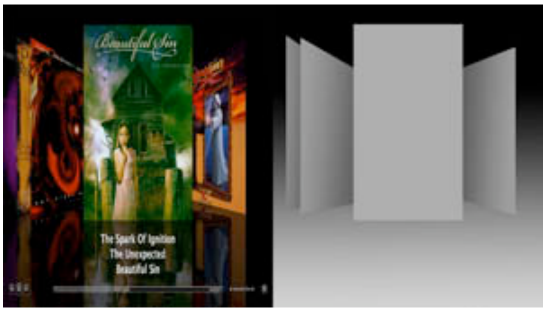


Figure 52: **Apple Cover Flow 3DUI**

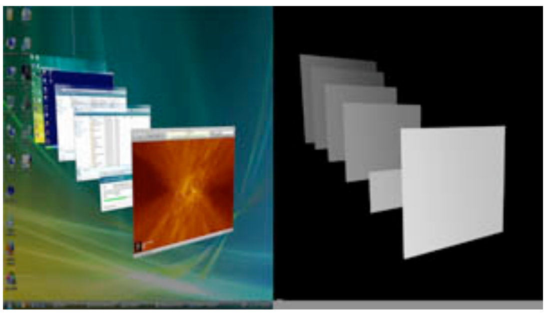


Figure 53: **Microsoft Vista 3DUI**

Interview

The questions, which were used in the interviews, can be divided into three categories. First, questions about 3D and 3DTV's, secondly about user interfaces and 3DUI's, and lastly into question about different forms of interaction.

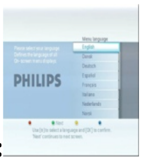
Questions - 3DTV

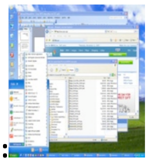
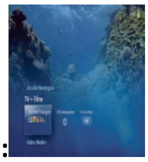
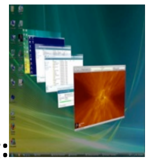
1. What do you know about 3D, have you seen it in Theme Parks or in Cinemas?
2. What do you expect of 3D in the future?
3. Have you ever heard of 3DTV?
4. Do you think 3DTV has advantages over a 2DTV?
5. Do you think 3DTV has disadvantages over a 2DTV?

Questions - 3DUI

6. Do you understand the principle of a User Interface (or menu)?
7. Do you know any 3D User Interfaces or 3D menus?
8. Do UI's such as in Windows XP look as 3DUI's to you?
9. Would a User Interface in 3D (a 3D menu) have advantages over plain 2D menus?
10. Would a 3DUI have disadvantages over 2DUI?

Look at the following UI's:

a: 

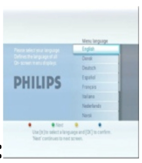
b: 
c: 
d: 
e: 
f: 
g: 

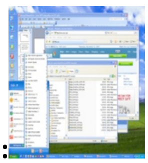
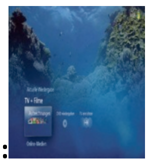
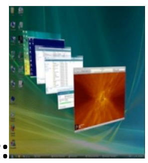
11. How would these examples look in 3D (if shown on a 3DTV)?
12. What could be the advantage if these UI's would be in 3D?

Questions – Interaction (gestures)

13. Do you understand the principle of interacting with media devices?
14. What forms of interaction with media devices do you know?
    - Interaction with TV, DVD players, pc's, game consoles or MP3 players.
15. Do you understand the principle of gestures as an interaction form?
16. Do you know gesture like forms of interaction?
    - One handed or two motion, pointing, waving, or gestures.
17. What kind of gestures do you think you would use with a 3DTV?

Look at the following UI's again:

a: 

b: 
c: 
d: 
e: 
f: 
g: 

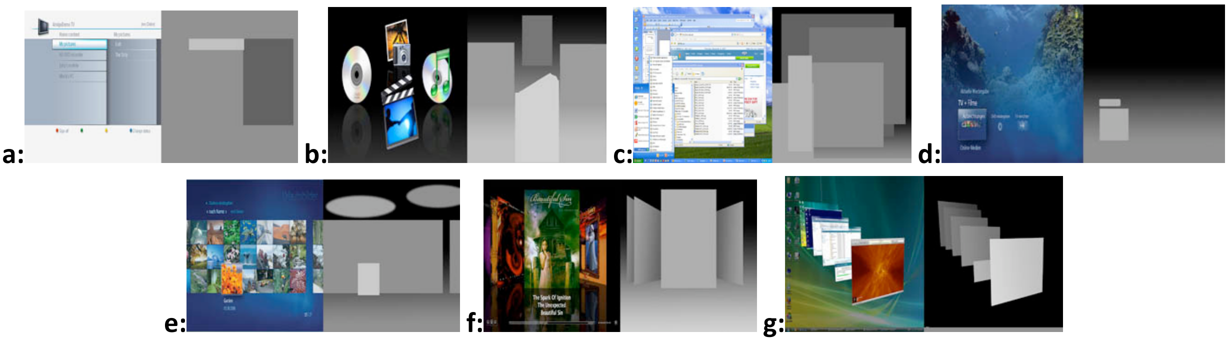
18. How would you use gestures to interact with these examples?

QUESTIONNAIRE

For the questionnaire, participants had to rate the 3DUI examples on the following three statements.:

- *The 3D effect is very clear!*
- *The 3D effect makes the UI clearer!*
- *The 3DUI invites the use of gestures!*

The 3DUI examples used in the questionnaire:



Participants had to rate using a seven-point Likert scale, ranging from strongly disagree (1) to strongly agree (7), and four (4) being neutral. During these ratings, participants were asked to think-out-load so the interviewer could note down any interesting and relevant thoughts.

strongly disagree		moderately disagree		slightly disagree		neither disagree nor agree		slightly agree		moderately agree		strongly agree
1	–	2	–	3	–	4	–	5	–	6	–	7

B. Pilot Study: Interview Summary

This Appendix is divided in three sections, questions about 3DTV, 3DUI, and Interaction. All provide a description of the comments given by the user during the interview per question asked. First, an overall description of the most commonly mentioned answers is given. These descriptions are added with an overview and sum of the most commonly answered topics (in light gray). And secondly, the results of each of the seven UI example screenshots will be discussed.

3DTV

EXPECTATIONS OF 3D IN THE FUTURE

When the participants were asked about their expectations of 3D in the future, the most common mentioned answers were that they expected 3D to venture into the home environment, into movies, but in particular into gaming applications. Most participants expect an increase in experience, both viewing and gaming, though almost all participants emphasized that they did not want to compromise in resolution or quality, compared to existing 2D standards.

In the home environment 2x
Gaming 2x
Minority Report like
That the PQ will be the same as 2D (full HD)
That there will be much more content created for 3D (Hype)
More experience
New way of looking at video
3D via broadcast (or via internet)
3D on blue-ray and HD-DVD's
2D to 3D switch-able display

ADVANTAGES OF 3DTV'S

When asked what the main advantage of a 3DTV over a 2DTV was, most participants mentioned that the extra dimension could result in a more immersive experience. Again the advantage in the gaming applications was occasionally mentioned. One participant nicely compared it to the switch from black & white to color TV's, adding 3D to the TV would simple add more realism to the viewing experience.

"Like the switch from black & white TV to colors, adding more realism to it!"

More immersive 2x
To lose yourself in a game or movie
Games 2x
Not looking at a character, but you are! (First person instead of 3rd person)
Look around a corner
Extra dimension 4x
More realism (like the switch from black & wit to color)
More experience 5x
More information
Special effects, like fright and scare moments.

DISADVANTAGES OF 3DTV'S

The most common mentioned disadvantages were picture quality on the one side, and eyestrain, head ages, and dizziness on the other side. Though, these answers sounded more as negative expectations which could result in real disadvantages. Another point which was often mentioned was that the content needed to be perfect otherwise one could easily be distracted, annoyed, or even irritated. This off course would then destroy your experience. One participant nicely raised the question if 3D could deliver the expectations that the hype had created, sounding as a form of doubt about 3D.

"Can it deliver the expectations?"

What might be strange is that, although participants knew that a lenticular 3DTV will be used in this project, meaning that no special goggles are needed to see 3D, still, some participants emphasized that using goggles must be seen as a disadvantage of 3DTV.

Comfort (how long can you look at it?)
Eyestrain 3x
PQ 3x
Everything needs to be right otherwise you will be easily annoyed. 3x
2D to 3D conversion
Blur
Goggles 2x
Can it deliver the expectations?
Viewing angle
Motion sickness
More different forms of DVD's (with 3D, without, hybrids), creates packaging problems.
Head ages and dizziness

3DUI

Familiarities with (3D) User Interfaces

Almost none of the participant knew real 3DUI's. Some though that the new Microsoft Vista contained 3DUI's, but none of them had actually seen the operating system in action. When asked if Microsoft XP had any 3DUI's, again, none of the participants confirmed. Microsoft XP was either seen as a stack of papers, or purely ecstatic 3D effects, also referring to the term 2.5D. They only resemblance some participants gave were in relation to games, or as one said, in a Virtual Reality (VR) setting, such as in CAVE.

Some within games 2x
Only those that suggest depth
Vista (I think) 2x
Vista 3D desktop 2x
Apple stacked items
Apple tv like
CAVE
No 3x

Advantages of 3DUI's

The question if 3DUI's would have advantages over plain 2DUI's where most commonly answered with a *"no"*. Most participants saw the added value merely as a fancier or nicer gadget. Only a small amount of the participants believed that the extra dimension, the third dimension, could make the UI more efficient, natural, or clearer. One of those participants framed this very nicely referring to the conceptual model of the UI.

"With a 3DUI you might create a better conceptual model of the interface's structure!"

Medical
Higher realism (Games)
Could create a better conceptual model of the UI
More efficient 2x
More information 2x
More natural (pushing a button metaphor) 2x
Clearer, Simpler representation 2x
Nicer, fancier 5x
Depth (extra dimension) 4x
Experience
Still see the parent menu
Less cluttered
More design freedom

Easier to navigate

DISADVANTAGES OF 3DUI'S

For the disadvantages of a 3DUI over a 2DUI, participants mainly mentioned problems as *"too much"* and *"unclear"*. With these expressions, participants mend that a 3DUI will probably be harder to design, and thus as a consequence it would be easier to screw-up if the design was not done properly. Another disadvantage that was mentioned, again, was having too use goggles for simple TV interaction such as volume changes etc.

"I don't want to put on goggles just to control my TV!"

You have to switch focus
Interaction problems 2x
The need for 3D interaction devices
To much – complex, (door de bomen het bos niet meer zien) 4x
Age
Strange (verwarrend)
Learning curve
Goggels to control the TV 2x
Expectations to high
No

EXAMPLE UI'S; HOW WOULD THIS UI LOOK ON A 3DTV?

The next Section will discuss the results of each the seven UI example screenshots, on the following question; how would this UI look if it was shown on a 3DTV?

Example 1: Philips Easy Logic TV UI; "displayed on a 3DTV"

Almost all participants envisioned how the Philips Easy Logic TV interface would look on a 3DTV in the same way. First, the selected (highlighted) item needs to pop-out of the screen like a button, and second, the gray area on the right must lie deeper into the screen. Only one participant mentioned that the selected (highlighted) item does not need to pop-out because it's only a highlight, not a button. Occasionally, participants mentioned that the image of the TV (top left corner) could also be in displayed in 3D, however, this had no direct relation to the real interface, but was purely an ecstatic remark.

Example 2: Apple Front Row UI; "displayed on a 3DTV"

Looking at the Apple Front Row interface, most participants agreed that the front (bottom) centered icon should come more out of the screen. Both left and right icons were on the same depth, and might come out of the screen only a little, whilst the back (top) icon should lie more into the screen, further into the depth.

There is a big difference between Apple and non-Apple users. Apple users immediately recognize this interface as a carrousel that can move in a circle, making it much clearer that the design, sort-of, is already in 3D. However, some non-Apple users just see this interface as four flat icons, no relation to a circular carrousel what so ever, nevertheless, they still give the same description based on the front icon being bigger then the back icon.

Example 3: Windows XP UI; "displayed on a 3DTV"

Half of the participants skipped this question as they saw no use or added value of creating Windows XP in 3D. Those who tried to envision, all came up with the same description. The active screen lies in front, coming out of the screen the furthest, then in the opposite way, the background (desktop) lies totally in the back, farthest into the depth. All the other (inactive) screens are divided equally in depth, between the front screen and the background.

Example 4: Vista Media Centre UI; "displayed on a 3DTV"

With the Vista Media Centre interface, all participants made a clear separation between the content in the background and the interface hovering in front. They all described it as if the content was pushed into the screen. Looking more closer to the UI, the highlighted button then pops out of the screen as the rest of the UI stays on the same depth as the actual screen.

Some of the participants mentioned that the background, the content, should be in 3D as well. This, off course depends on the type of content and is, therefore, unrelated to how the interface should look when displayed on a 3DTV.

Example 5: Media Centre Image Viewer; "displayed on a 3DTV"

On the Vista Media Centre Image Viewer interface, again, most participants agreed on how this would look if it were to be displayed on a 3DTV. The whole row of pictures needs to come out of the screen a little, and the highlighted button needs to come out of the screen the most. The text needed to be on the same level as the pictures, and the background lies farthest into the depth, behind all the other things.

Example 6: Apple Cover Flow UI; "displayed on a 3DTV"

In the following example, most the participants were very clear about how this interface would look if it would be displayed on a 3DTV. The centered CD pops out of the screen, with all other CD's stacked in depth behind it. For the first time, participants mentioned that objects, the outer CD's, where laying obliquely into depth. The text and status bar should be all the way upfront, Although some participant doubted if the text should be on the same level as the centered CD, yet this would then given problems with the reflection of the CD cover on the table.

Example 7: Microsoft Vista; "displayed on a 3DTV"

For the final example, the Microsoft Vista interface, most participants found it an easy and logic interface to envision on a 3DTV. There should be a clear difference between the screens, popping out of the display, and the background, lying totally in the back. They other screens should be equally divided amongst the front one, and the background.

INTERACTION

The third part discusses people's interaction familiarity with different media devices. More specifically, it categorizes gesture interaction into different groups and discusses people's preferences. Again, this part also shows the seven UI example screenshots, but now `participants have to visualize that the UI is in 3D and shown on a 3DTV. They are asked how they would interact with these UI's, given that it would have to be some form of gestures as interaction style.

FAMILIARITY WITH MEDIA INTERACTION

The most commonly used devices and interaction forms are keyboard and mouse, the remote control, and the physical buttons on a device (such as MP3 players). These are followed, but in a lesser extent, by your typical game controllers from the playstation, Xbox of Nintendo. And lastly, the mouse touchpad of a laptop, and the touch screens of PDA's.

Interaction forms that were only occasionally mentioned are game console extras (such as the iToy, force feedback, wheel and pedals, and guitar etc.), voice and speech control, face and eye tracking, and pointing devices or wands (Gyroscopic).

Strangely, especially because the purpose of this interview was explained to all participants at the beginning, gesture interaction was only mentioned twice.

Categorization of Gesture Interaction

When moving more in specifics towards gestures as an interaction style, and after explaining some more about it, the best way to categorize gesture interaction resulted in the following groups:

- Point and Click (if accurate enough)
- Small Hand-Gesture Recognition (characters - like alphabet with fingers)
- Gesture Recognition (signs and symbols)
- Moving Gesture Recognition (with arm movements)
- Manipulation (of virtual objects)
- Identification (personal identification)
- Multiple Hand Interaction

Although not all participants agreed with this specific categorization, looking at a whole, these were the most common mentioned groups.

Preference of Gestures in combination with a 3DTV

When the participants were asked what the best way is to use gestures in combination with a 3DTV, all (most) participants said to keep it basic and simple.

"Not to big, it must be as easy as the remote control is now"

Most people could imagine using their hands as point-and-click devices, but all emphasized that if gestures would be used, feedback would be most important. If there would be sufficient, correct, and real-time feedback, it might be possible to use manipulation as well. What might be remarkable is the fact the one particular movie (with Tom Cruise, 2002) kept popping-up,

"Like in that movie, Minority Report!"

Example UI's; how would you Interact with Gestures?

The next Section will discuss the results of each the seven example UI screenshots, on the following question; how would you interact with the UI if it would be shown on a 3DTV, and given that you would have to use gestures as a way of interacting.

Example 1: Philips Easy Logic TV UI; "interacting with gestures"

Most commonly mentioned type of gesture interaction with the Philips Easy Logic interface was point and click, though several participants found it useless to use gesture interaction with this UI.

Example 2: Apple Front Row UI; "interacting with gestures"

With imaging gesture interaction with the Apple Front Row interface, again, a clear difference between participants with prior Apple experience, and those without, was experienced. The participants, whom where familiar with this form of interface, immediately envisioned a hand swing to turn the carrousel around. However, Non-Apple users, who did not saw the principle of the carrousel, choose point and click as a interaction style. In other words, they envisioned themselves simply pointing at on one of the four icons, and clicking to activate.

Unexpectedly, some participants envisioned themselves grabbing the front icon and pulling it towards themselves to activate, instead of pushing it to activate as you would normally do with a mouse click.

Example 3: Windows XP UI; "interacting with gestures"

For most participants, 3D windows XP would not invite any form of gesture interacted. There was only one participant who envisioned grabbing forms and pulling them to the front. But, all in all, almost all dislike this idea.

Example 4: Vista Media Centre UI; "interacting with gestures"

For this example, all participants mentioned simple point and click as the preferred type of gesture interaction. For them, the button was so clear that there was nothing else to do instead of pressing it. Although, the button was obvious, al lot of participants were not invited to do other things with the interface.

Example 5: Media Centre Image Viewer; "interacting with gestures"

As with the former example, the most commonly mentioned type of gesture interaction for the Media Centre Image Viewer was point and click. Moreover, this time the participants were invited to do more, apart from clicking the highlighted picture, they envisioned themselves pointed their finger for selecting other pictures.

Example 6: Apple Cover Flow UI; "interacting with gestures"

How to interact using gestures with the Apple Cover Flow interface was very clear for most participants. With a swipe of the arm you could swap to another CD, and clicking on it would then activate it. Some would even want to grab the CD and turn it around to see the backside.

Example 7: Microsoft Vista; "interacting with gestures"

For the last example, the Microsoft Vista interface, the most common mentioned gesture interaction style was point and click. Additionally, some participants thought about pushing windows away. Though this UI invites the use of gestures with most participants, several participants emphasized that it would only work if there was a clear form of feedback.

C. Pilot Study: Questionnaire Raw Data

This Appendix is divided in two sections, the raw data of the questionnaire and a list of the most preferred interfaces per question and overall.

Raw Data

	User Interface																				
Participants	a			b			c			d			e			f			g		
1	6	4	5	5	5	3	7	6	2	5	2	4	6	3	2	6	5	5	7	6	6
2	7	4	3	6	5	5	7	6	3	7	7	4	7	7	4	7	6	6	7	4	3
3	7	6	4	7	7	6	7	7	4	6	5	5	7	7	5	7	7	7	6	5	4
4	2	1	2	5	5	6	2	1	1	4	4	5	6	4	6	4	4	5	6	6	7
5	5	4	5	6	6	6	7	3	3	6	6	6	7	6	6	6	7	7	6	5	6
6	5	4	5	4	3	3	4	5	3	5	5	4	5	5	5	5	3	3	6	4	6
7	5	2	1	6	6	6	6	5	3	7	6	2	6	7	6	7	7	7	7	4	5
8	6	5	4	5	4	7	7	6	6	6	6	6	7	6	6	7	7	7	5	6	6
9	3	2	4	5	5	7	6	6	1	6	4	5	6	4	5	6	5	7	6	6	5
10	7	5	6	5	5	6	7	7	7	7	5	5	7	6	7	6	5	5	7	6	5
11	5	5	5	4	5	6	7	5	6	6	6	7	6	6	4	7	6	7	7	6	7
12	5	2	2	6	5	6	6	5	5	7	3	4	6	6	2	6	6	6	6	6	6
13	3	4	2	4	4	4	5	4	2	6	5	7	6	5	6	5	5	6	7	6	6
14	4	5	1	5	5	4	7	2	5	6	4	7	6	4	5	6	4	6	6	3	2
15	6	5	5	5	5	6	7	7	7	7	5	5	7	5	5	7	6	6	7	6	7

List of most preferred UI

Question 1: The 3D effect is very clear!

1. Example 7: Microsoft Vista 3DUI (M = 6.40, SD = .63)
2. Example 5: Media Centre Image Viewer 3DUI (M = 6.33, SD = .62)
3. Example 6: Apple Cover Flow 3DUI (M = 6.13, SD = .92)
4. Example 3: Windows XP 3DUI (M = 6.13, SD = 1.46)
5. Example 4: Media Centre 3DUI (M = 6.07, SD = .88)
6. Example 2: Apple Front Row 3DUI (M = 5.20, SD = .86)
7. Example 1: Philips Easy Logic TV 3DUI (M = 5.07, SD = 1.54)

Question 2: The 3D effect makes the UI clearer!

1. Example 6: Apple Cover Flow 3DUI (M = 5.53, SD = 1.25)
2. Example 5: Media Centre Image Viewer 3DUI (M = 5.40, SD = 1.24)
3. Example 7: Microsoft Vista 3DUI (M = 5.27, SD = 1.03)
4. Example 2: Apple Front Row 3DUI (M = 5.00, SD = .93)

5. Example 3: Windows XP 3DUI (M = 5.00, SD = 1.18)
6. Example 4: Media Centre 3DUI (M = 4.87, SD = 1.30)
7. Example 1: Philips Easy Logic TV 3DUI (M = 3.87, SD = 1.48)

Question 3: The 3DUI invites the use of gestures!

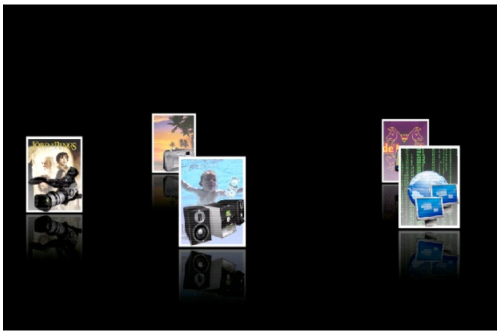
1. Example 6: Apple Cover Flow 3DUI (M = 6.00, SD = 1.13)
2. Example 2: Apple Front Row 3DUI (M = 5.40, SD = 1.30)
3. Example 7: Microsoft Vista 3DUI (M = 5.40, SD = 1.45)
4. Example 4: Media Centre 3DUI (M = 5.07, SD = 1.39)
5. Example 5: Media Centre Image Viewer 3DUI (M = 4.93, SD = 1.44)
6. Example 3: Windows XP 3DUI (M = 3.87, SD = 2.03)
7. Example 1: Philips Easy Logic TV 3DUI (M = 3.60, SD = 1.64)

Overall UI score: The three questions combined

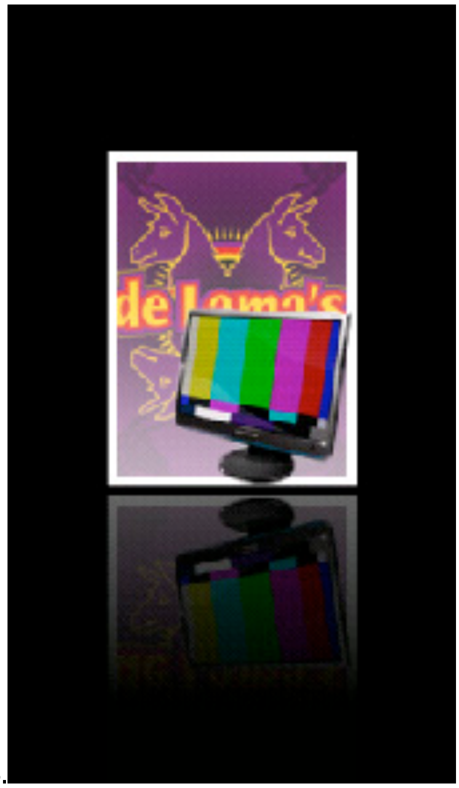
1. Example 6: Apple Cover Flow 3DUI (M = 6.00, SD = 1.13)
2. Example 7: Microsoft Vista 3DUI (M = 5.69, SD = .81)
3. Example 2: Apple Front Row 3DUI (M = 5.40, SD = 1.30)
4. Example 5: Media Centre Image Viewer 3DUI (M = 5.33, SD = .76)
5. Example 2: Apple Front Row 3DUI (M = 5.20, SD = .84)
6. Example 3: Windows XP 3DUI (M = 5.00, SD = 1.45)
7. Example 1: Philips Easy Logic TV 3DUI (M = 4.18, SD = 1.28)

D. Main Study: Design of the Interface

The design of the interface (1) was based on the Apple Front Row interface (2). Icons are dislayed on an oval circle, which suggest that it could turn round like a carrousel. For the experiment, a carrousel with five icon was created. Each icon represented a menu item of a media center like application. For example, listening to music (3), watching a movie (4), showing foto's (5), watching TV (6), or going on the internet (7).

1. 
2. 
3. 
4. 

5. 

6. 

7. 

E. Main Study: Demographic Questionnaire

1 ***Participant number:*** *…………………………………………………*

2 ***Date:*** *………………………………………………………………………*

3 ***Time:*** *………………………………………………………………………*

4 ***"Please enter your name:"*** *………………………………………*

5 ***"What is your sex?"*** (Male / Female)

6 ***"What is your age?"*** *……………………………………*

7 ***"What is the highest level of education you have completed?"***
(High school / Bachelor / Master / PhD)

8 ***"Do you think of yourself as technical?"*** (Technical / None technical)

9 ***"Are you an Apple* user?"*** (Yes / No)

**With an Apple user we mean if you are familiar the Apple products such as the iMac of iBook, and with the Apple Operating System (Tiger OS X or Leopard), familiar with using iTunes and iPod music player, or familiar with an iPhone?*

10 ***"Are you right or left handed?"*** (Right / Left)

F. Main Study: Informed Consent From

SUBJECT: Easy 3DTV Interaction

____

Name Participant

Every participant needs to sign an informed consent form before participating in the experiment. By this they declare to understand what is expected from them and what their rights are.

I voluntarily agree to participate in the experiment about "Easy 3DTV Interaction". I understand that this experiment and interview is being conducted by Berrie Duijx, the Program Director, to improve the interaction of 3DTV's and is also a part of his graduation project.

I understand that the evaluation methods, which may involve me, are:

1. The Program Director's recorded observation of the experiment and/or;
2. My participation in a 45-minute interview.

I grant permission for the experiment and interview to be recorded transcribed, and to be used only by Berrie Duijx for analysis of experiment and interview data. I grant permission for the evaluation data generated from the above methods to be published in an evaluation report to the Eindhoven Technical University, Philips CL iLab, and in the dissertation and future publication(s).

I understand that any identifiable information in regard to my name and/or agency name may be listed *only* in the above-mentioned evaluation report to the Eindhoven Technical University, that is, this information will *not* be listed in the dissertation or any future publication(s).

Hereby I give permission to record a video of my experiment. I understand that this information will not be distributed outside the project.

Optional:

- I give permission to present the video of my experiment in public (only small 15 second fragments for educational or research purposes, to give some gesture impressions): Yes / No
- Optional: I want to be on the email list for future experiments invitations? Yes / No

______________________________	_________________________________
Signature Research Participant	Email Adres Participant

____

Date

G. Main Study: Frame of Reference

Quality of the 3DTV:

Please keep in mind that future versions of such a 3DTV will be of better quality. For example, one could expect a higher resolution and more viewing angles, therefore, do not score the following questionnaire by judging the quality of this particular 3DTV.

Environment:

Please keep in mind that this experimental setup will be tested for the use in a living room (a televised home entertainment environment). Therefore, picture yourself sitting in a living room on a couch, approximately 3 to 4 meters away from the TV.

Pointing device:

With a pointing device we mean a device, which lets you interact with the TV by pointing at or towards it.

Gestures:

With gestures we mean hand and arm gesture movements, which could manipulation of virtual objects, in other words, simulating the intuitive manipulation of everyday objects. We do not mean that you will have to learn or know a hand-sign language or such.

Carrousel like interface:

Do you understand the principle of a carrousel like interface? Take the following Apple Front Row for example:


Photos

H. Main Study: Quantitative Questionnaire

"The 3D effect is very clear"

strongly disagree		moderately disagree		slightly disagree		neither disagree nor agree		slightly agree		moderately agree		strongly agree
1	–	2	–	3	–	4	–	5	–	6	–	7

"The 3D effect makes the UI clearer*"

**Do you think the 3D effect has added value to the UI?*

strongly disagree		moderately disagree		slightly disagree		neither disagree nor agree		slightly agree		moderately agree		strongly agree
1	–	2	–	3	–	4	–	5	–	6	–	7

"The UI invites the use of gestures"

strongly disagree		moderately disagree		slightly disagree		neither disagree nor agree		slightly agree		moderately agree		strongly agree
1	–	2	–	3	–	4	–	5	–	6	–	7

"I would dislike interacting with this UI using gestures"

strongly disagree		moderately disagree		slightly disagree		neither disagree nor agree		slightly agree		moderately agree		strongly agree
1	–	2	–	3	–	4	–	5	–	6	–	7

"I would prefer to interact* with this interface, using:"

**please keep in mind that this setting will be in a living room!*

(Not at all / Remote Control / Keyboard and Mouse / Pointing Device / Gestures)

"If you had to use gestures to interact, how would you then*:"

**Please think out loud and actually make yourself visual, really use your hands and arms!*

- How would you select or activate an item?
- How would you select another item?

I. Main Study: Interview Topics

Problem of unintended gesturing

When should the menu become active or react?

What to do with unintended interaction?

What do you want to do to ignore or cancel an unintended interaction?

Problem of controlling the interaction

Who should have control of the interaction? (In case of a TV with a remote control (RC), the one who has physical possession of the RC, controls the interaction with the TV).

Problem of energy costs

What do you think of constantly waving your arms?

Would you see this as a problem?

Would you get tired or irritated?

Problem of no tactile feedback

What do you think could serve as a solution for the fact that you have no tactile feedback on your interaction?

J. Main Study: Qualitative Questionnaire Raw Data

This Appendix is divided in four sections, questions about the problem of unintended gesturing, problem of controlling the interaction, problem of energy cost, and the problem of not having tactile feedback while gesturing. All provide an overview and sum of the most commonly answered topics given by the user during the interview per question asked.

Problem of unintended gesturing

- Eye track device, gaze control. Active when you watch/look at the TV 15x
  - Face recognition 3x
  - That is also a natural thing to do, like with humans 2x
  - No eye contact
- Voice control 10x
  - Code word
  - Activation command
- One hand gesture starts the menu 6x
  - Point finger at TV to activate 4x
  - Close hand first (or other gesture) 2x
  - Stretch hand 2x
  - Move towards the TV
  - Code gesture 2x
  - Flick of the finger 4x
  - Tap your knees
  - First a sign to get control
  - Like pointing the remote
- Learning system
  - When watching a movie its unlikely that gesture will occur
  - When gesturing for the 10x it's likely that a 11x will follow
- Could be very irritating, the technology must work
- Different ways of activating it.
- Same problems as with voice commands, multiple at the same time.
- Prevent unintended gesturing
- Can make some errors because it is new and innovative
  No long learning curve
- Different per application 2x (context sensitive)
- First activate then listen for gestures 2x
  - Activation should be a gesture that you normally wouldn't make 3x
  - No standard gestures
  - Brain
  - Blink
  - Build in something that lets the TV know, what follows is for me.
- Zero error tolerance
- When you want to
- A cancel gesture 4x
  - Cross

    - Through away 2x
    - One hand gesture
    - Form of undo, undo the last gesture
    - Flat hand
    - Shake the head
    - Undo features
- Children
- Not first a voice control
- ID recognition
    - Calibrate
    - Create your own gestures
- Like with the RC now
- In front, between the person's body and the TV
- Body posture
- After 15 seconds in the none gesture mode
    - Indication on TV if in gesture mode (LED)
- No voice
- No device
- Want to see what you are doing on the screen

Problem of controlling the interaction

- Social problem 18x
    - TV etiquettes
    - But with a referee, main controller
- Must react on everybody 12x
    - The one closed to the TV has control
    - Passing through the control, but with approval to the TV
    - Easy pass through of control
    - Pass through control… how?
    - You should not want that there is one boss
- Physical device 7x
    - Some sort of physical device that can be passed around, which enables the TV to see who has control
    - A ring or arm bracelet 3x
    - Red ball
    - Magic wand
    - With your remote in gesture mode
    - A remote control to switch gestures off 2x
        - Like during a movie 2x
    - Always a remote control
- Only one person should have control 5x
    - Calibrate on body temperature
    - Calibrating with facial camera (schiphol camera ID) 2x
    - The one that switches on the TV is the main user
    - The one that activated the menu 2x
    - Needs to know who paid for it
    - Different for the time of the day
    - Voice recognition 2x
    - Lock on yourself
- Some form of user configuration 6x

  - Dad
  - Boss
  - A main user
  - Master control
  - Configure power
  - Disable kids 4x
    - Everything below 140 is out
    - Size of finger tips
  - Form of parental control 2x
  - Face ID
  - Calibrate
- Individual teaching, learning system 3x
  - Scan for emotions
  - No control when negative emotions
- Straight in front of the TV
- Not react when someone shows ADHD behavior
- No idea
- Same problem as with a remote control
- Size of the hands
- Children have to learn how a future TV works
- So four remotes
- Different modes
- Clear feedback

Problem of energy cost

- Not a problem 22x
  - People are used to making gestures
  - Doesn't get used constantly
  - However, your elbow must be supported 3x
    - Only wrist and hand are ok, total arm movement are not.
  - If Intuitive 2x
  - If application is content orientated that there is no problem. Fun!
  - But it should not become a Wii training
  - Searching the remote costs energy to 2x
  - Even better, do not need to get up to get the RC 2x
  - Even better, do not need to search the RC
- UI design problem 3x
  - Depends on how the menu is built up
  - Could be while gaming
- Yes 3x
  - Especially with simple things like volume and channel select
  - Will always be a problem with gestures
- Maybe
  - Minimal like with zapping 3x
  - Minimal 3x
  - Limit to finger or wrist
  - Fun WOW effect will diminish after a while
  - Personalize
  - Both, gesture + RC
- Psychological energy 3x

    - Dutch house with open curtains. The neighbors see you doing strange things with your arms… “ That man is ill”
- Learning system
    - First big movement, later smaller
- Option for older people or disabled like Parkinson

Problem of no tactile feedback

- Should not be a problem with good recognition and feedback 20x
    - Like an ipod, you also do not feel any resistance, but act on the visual and additive cues
    - RC has no feedback either
- Good feedback 4x
    - Visual feedback 12x
        - For example icon, highlight 2x
        - Display something on the screen where you are
        - Want to see that the TV is listening to gestures. Like with MSN, you can see that it is typing
    - Feedback must be direct 9x
    - Feedback through sound 8x
    - Led light 3x
    - You expect feedback
- Needs to have a learning curve 4x
- Yes 3x
    - RC you can feel that you press
    - Both, the new UI and the old UI

K. Main Study: Overall Reactions

This Appendix provides an overview and sum of the most commonly mentioned reactions and comment made during or at the end of the experiment.

REACTIONS

- No reactions 18x
- Technical problem 3x
    - Afraid that the movement must be to big
    - Afraid that it will only work in front of the TV but not while walking around, in the living room
    - Gestures would work when only one meter away, not when sitting on the couch.
- Only if it would work 100% 2x
    - Must work the first time
- 3D triggers more gesturing 2x
    - You want to grab behind something
- I would want! Costs?
- You are triggered to sweep everything to the right because the UI is turning to the right.
- Got the feeling that you can pull it more to the front.
- No WOW without 3D binocular cues
- Something that follows your hand
    - Pointer on the screen
    - Or a virtual grip on the screen, suggesting gesturing
- Comments on the UI design
    - Highlight the selected
    - Make smaller faster
    - Not to much overlap

  Motion was an added value
- Could be more interesting if icons where not flat
- 3D effect is irritating
    - Head ach
- Could not recognize a carrousel in the UI, maybe with more icons
- Maybe start with a RC with pointing
- Display a Red blinking led when gesturing, as when using the RC
- The movement creates stress. I need to be on time to press otherwise I need to wait a full lap.
- Movement helps to understand that you can move the UI
- Pupil tracking
    - Interaction with where you look at
- Prefer to gesture standing up than sitting down.
- Viewing angle of the camera problem
- Good thing, no RC. (Use a function on your GSM!?)
    - No buttons, not lost, not broken
- What about disabled
- Is it a Need or a Gadget?
- For see large problems with for example the fixed menu's of DVDs etc.

L. Definitions

“Easy interactions”

“Easy Interactions” is an ITEA 2 project. ITEA 2, the follow-up to the successful ITEA program, is a strategic pan-European program for advanced pre-competitive R&D in software for Software-intensive Systems and Services (ITEA2, 2008).

The Easy Interactions project proposes innovative approaches to human-system interactions (HSI) for application domains that require enhanced HSI, such as critical situation, mobility, business, home or industry. Enhanced HSI will not be addressed by the juxtaposition of different interfaces but through their combination; the technologies studied within the project are therefore multimodality, augmented reality and context/user awareness. Implementation issues, interoperability and upgradeability will be addressed through an open architecture. Easy Interactions will apply a specification, implementation and validation process to the interfaces derived from the requirements of the targeted application domains, from scenarios up to demonstration (ITEA2, 2008).

“Easy TV interactions”

“Easy TV interactions” is a SenterNovem IS project. SenterNovem is an agency of the Dutch Ministry of Economic Affairs that promotes sustainable development and innovation, both within the Netherlands and abroad. They aim to achieve tangible results that have a positive effect on the economy and on society as a whole (Senternovem, 2008).

The project aims at the design, prototype and validate for gesture-based 3DTV-control making use of virtual objects (buttons, sliders, …) displayed by a 3DTV.

This IS project proposal is part of a much larger European ITEA2 project, with international partners that are eminently suited to realize the required material and to demonstrate their applicability in the 3DTV home environment.

Philips Consumer Lifestyle Innovation Lab

Philips Consumer Lifestyle (Philips, 2008) is a world player in the world of consumer electronics, domestic appliances, and personal care products. The international Know-how Centre of the consumer electronics division of Philips Consumer Lifestyle (CL) is the Philips Innovation Lab (iLab). The iLab provides and drives new features and concepts, creating innovations that could have substantial positive business impact for the Philips Consumer Lifestyle sector. Philips, and so the iLab, support the brand-promise "Sense & Simplicity" by being designed around the customer and by being easy to experience.

Philips 3D Solutions

3D Solutions (Philips, 2008), currently a Philips Technology Incubator, brings the 3D viewing experience to the viewer without the need for special 3D glasses. 3D Solutions deploys the displays and content creation tools to establish this three-dimensional (3D) viewing experience. Technology Incubators help to identify new business opportunities and transform ideas into new and successful businesses. 3D displays from 3D solutions are currently being deployed in the Professional domain, and are being used for digital signage and point of sales applications.

Royal Philips Electronics

Royal Philips Electronics of the Netherlands is a global leader in healthcare, lighting and consumer lifestyle, delivering people-centric, innovative products, services and solutions through the brand promise of "sense and simplicity". Headquartered in the Netherlands, Philips employs approximately 123,800 employees in more than 60 countries worldwide. With sales of EUR 27 billion in 2007, the company is a market leader in medical diagnostic imaging and patient monitoring systems, energy efficient lighting solutions, as well as lifestyle solutions for personal wellbeing. News from Philips is located at http://www.philips.com/newscenter.