Kari-Jouko Räihä, Kimmo Koivunen, Harri Rantala, Selina Sharmin, Turkka Keinonen, Sami Kukkonen and Sami Lahtinen

Perception of Design
Tekes project 2003-2005
Final report

DEPARTMENT OF COMPUTER SCIENCES
UNIVERSITY OF TAMPERE

D-2006-4

TAMPERE 2006

UNIVERSITY OF TAMPERE
DEPARTMENT OF COMPUTER SCIENCES
SERIES OF PUBLICATIONS D – NET PUBLICATIONS
D-2006-4, MARCH 2006

Kari-Jouko Räihä, Kimmo Koivunen, Harri Rantala, Selina Sharmin, Turkka Keinonen, Sami Kukkonen and Sami Lahtinen

Perception of Design
Tekes project 2003-2005
Final report

DEPARTMENT OF COMPUTER SCIENCES
FIN-33014 UNIVERSITY OF TAMPERE

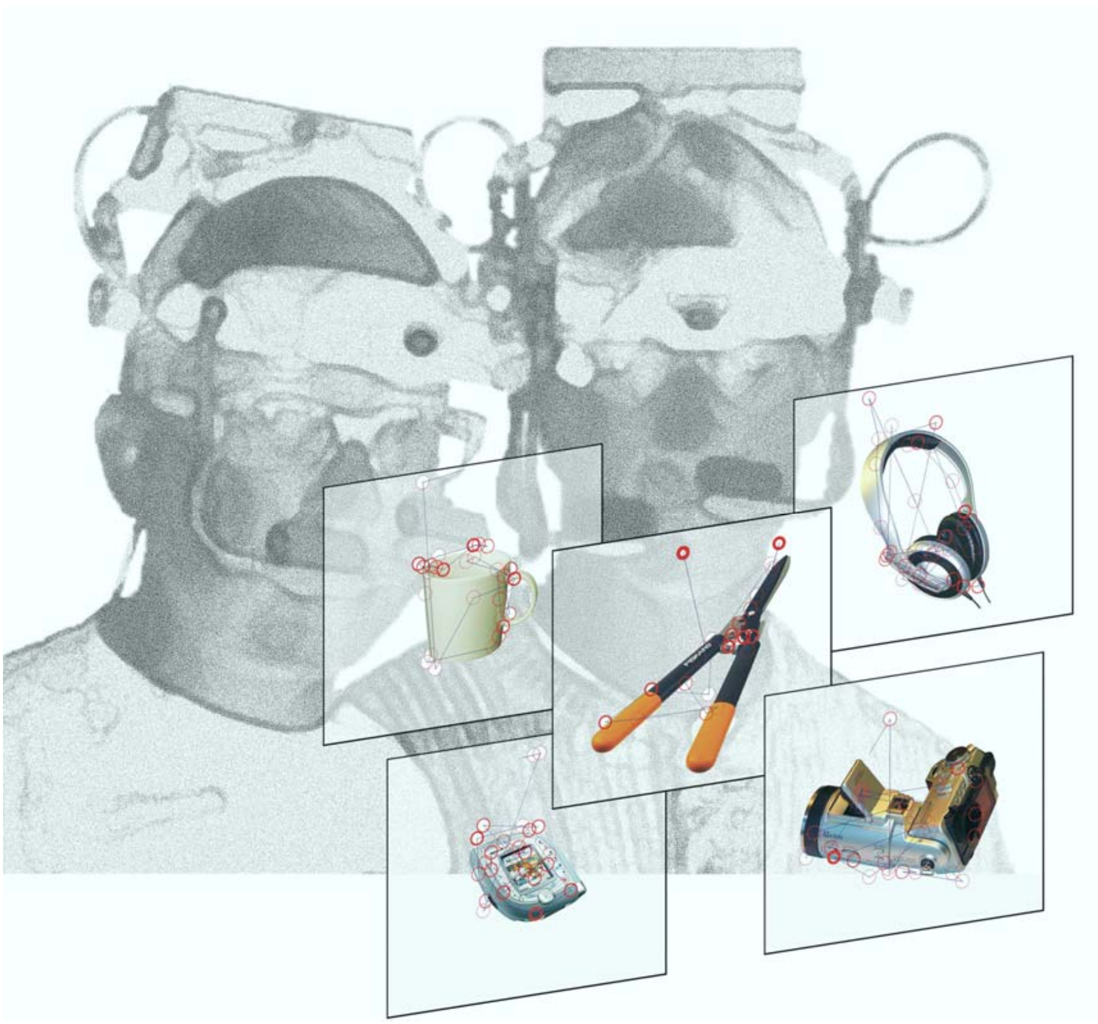
ISBN 951-44-6598-9
ISSN 1795-4274

perception of design

Tekes project 2003 – 2005, Final report

University of Tampere | Department of Computer Sciences
Kari-Jouko Räihä, Kimmo Koivunen, Harri Rantala, Selina Sharmin

University of Art and Design Helsinki | School of Design
Turkka Keinonen, Sami Kukkonen, Sami Lahtinen



DESIGN 2005

Contents

1 Introduction

In design research and in the education of designers, a holistic approach is prevalent. Since design quality is often understood through the relationship between a design product and its context of use, the details of a design product do not always get the attention they deserve. Managing the details and their interrelationships is, however, essential as much of the design process in practice consists of finalizing the details, choosing the best shapes and selecting the most appropriate colours.

This trend is partly caused by the lack of proper methods for studying the details of design products. Typical consumer research methods rely on consumers' opinions, which by their nature are inaccurate and depends on interpretation.

However, eye-tracking provides useful methods for evaluating the details of design for usability. This research project aimed at applying these established methods in the field of design research, in order to create an evaluation approach that is both feasible and reliable. The research was carried out jointly by University of Tampere having expertise in usability evaluation and eye-tracking research and University of Art and Design Helsinki having expertise in industrial design. Nokia, Fiskars and Clothing+ acted as business partners in the research project. This project was a part of the Design 2005 program organised by the Finnish National Technology Agency Tekes.

1.1 Aims and objectives

The project had three distinct but related goals:

1. Development of a research method for studying the perception of design products.
2. Comparison of how designers and non-designers[1] perceive design products.
3. Comparison of perception of 3D design products and their computerized 3D models.

On a general level, the aim was thus to develop a new set of methods, to apply it in experimental studies, to produce new knowledge on the perception of design products, and to integrate that knowledge in the education of designers. The results should be directly applicable in product development processes. We will discuss the three goals one at a time.

1.1.1 Development of a research method

Eye-tracking is utilized in several applications and research fields. One such field is the study of how people perceive physical objects. There is a connection between the eye and the hand movements—gaze precedes action. Eye gaze direction shows where our attention is directed. However, the information about the gaze direction alone is not enough without the knowledge of the features of the object and the intention of the observer. Thus, we need to combine eye-tracking with other methods of investigation. We need methods to distinguish the features of the product that capture the consumer's attention. Such features are, for example, the color or the shape of the object.

Which visual features of a product capture the observer's attention? Previous research shows that people do not randomly explore an image; for example, the items in the foreground get more attention than items in the background (Babcock et al., 2002). People attend to certain distinct features such as edges, asymmetries, contrast shift, or bold colours. Information of the

[1] In this report we use the term "designers" to denote people that have received their education in a design school, and "non-designers" to denote those who do not have such a background.

gaze behaviour, combined with the knowledge of the features of the product, reveals the features that capture attention.

Also motivation drives attention. People use visual information that is relevant to the task. A classical example of how the task affects the gaze path is the study by Yarbus (1965). The observer's gaze path changes depending on the given instructions.

A significant part of the visual information is processed on a preattentional level. The preattentively processed features only become apparent to the user if the features do not correspond with the mental representation of the features. For example, the shadow of an object is not perceived on a conscious level but it still significantly affects the perception of an object (Rensink and Cavanagh, 1993). The consumer may not be able to verbalize why one product is more appealing than another.

In this project we have continued the study of visual attention and focused on the perception of design products. Our goal has been to develop a method for studying perception of design products by combining (1) eye-tracking and gaze-path analysis, (2) think-aloud protocols in experimental tasks, and (3) comparative controlled studies of products that are manipulated so that they differ only on a given detail.

1.1.2 Comparison of how design designers and non-designers perceive design

Our second goal has been to produce new knowledge on the differences in the perception of design products by designers on one hand, and non-designers on the other hand. This is essential for understanding the foundations of user-centred design. For instance, if it is found that designers use a richer set of techniques than non-designers, this should be taken into account both in the design process and evaluation.

It is also interesting to study whether design is evaluated based on aesthetics, functionality, form, architecture, brand, or some other dimension—and whether designers and non-designers differ in this respect.

Similar expert/novice comparisons have been carried by out using eye-tracking in other domains, including surgery (Tchalenko, 2001), inspection of broken devices (Lu et al., 2001) and art (Miall and Tchalenko, 2001; Wooding, 2002). For instance, John Tchalenko has in his studies observed that artists perceive works of art differently from novices, and this observation has subsequently been used in art education.

1.1.3 Comparison of perception of 3D design products and their computerized 3D models

Eye-tracking has previously been used in studying printed and electronic media (Lewenstein et al., 1999) and in the context of two-dimensional user interfaces (Babcock et al., 2002; Jacob and Karn, 2003). Studying 3D objects using eye-tracking is a new technological challenge. In addition to 3D physical objects our goal was to study their 3D CAD models and virtual models. This is motivated by the important role that such models have in everyday design tasks. Remote usability testing of models over the internet is also becoming more common, and this requires methodological development (e.g. Battarbee et al., 1999; Kuutti et al., 2001).

Again, in addition to the practical design and evaluation implications, the problem is interesting in its own right: how does the perception of a physical 3D object differ from the perception of its digital model?

2 Background

2.1 Design evaluation

Several methods exist for evaluating product design for usability, sales, profit, and ease of manufacturing. Some of these methods deal with so-called extrinsic properties (added properties like price, brand, and promotion) which are outside of a designers' direct influence. Their opposite, the intrinsic properties, are the physical product properties which a designer can directly affect on. This study revolves mostly around the intrinsic product attributes.

Design is here understood as the visible features of a product created by the means of industrial design, not in the broader sense of the word including meanings like engineering or software design. The influences of various product attributes affecting the product's general appeal have been examined widely, but the relevance of visual properties has received less attention.

Design evaluation is affected by (1) the evaluator, (2) the purpose and nature of the evaluation, and (3) the characteristics of the object being evaluated. First, though there are certainly many meaningful ways to characterize and group the evaluators, in our studies we will use the simple categories "designers" and "non-designers". Second, there are many different ways to carry out the evaluation process. Evaluation can be quantitative or qualitative. Depending on when in the design process evaluation takes place, it can be formative or summative. The evaluation process can be based on the evaluator's strong attachment to previous evaluations, which has been called "affect referral heuristics" (Keinonen, 1998), or it can be a more systematic attribute based evaluation, an attribute being an aspect of a product that can be used for comparison. Third, evaluation is affected by the object of evaluation and its representation. Evaluation can deal with the activity around the object or concentrate on the evaluator's attitudes and preferences. The object can be physical or a visual two dimensional (2D) or three dimensional (3D) representation of the actual object.

With the categorization above, a usability test, for example, could be described as involving users in an activity of using a product for a detailed task. The analysis would be quantitative (based on performance time, for instance) and diagnostic by nature. Accordingly, the evaluation in this project can be described as one where both designers and non-designers carry out a subjective evaluation of both the details and the product as a whole, based on their attitudes and preferences. The analysis combines both the qualitative and quantitative data in a summative manner and concentrates on attribute based evaluation.

The means of product evaluation in this study include various tasks and viewing motivations. For example, the studied objects were (1) viewed freely without any particular tasks, (2) compared to other products in the same product category, (3) rated on a numeric scale, or (4) ranked according to personal preference. The tasks in categories (3) and (4) were mainly used to provide data that helps in interpreting the gaze data, which tells what the user is perceiving, but does not tell why the particular object or its attribute has attracted attention.

The objects in our studies can be categorized according to fidelity and the type of visual representation. Fidelity can be divided to ideas, concepts, and final or almost final products. When these are combined with the types of representation, we end up with 2D sketches, concept drawings, photo realistic renderings and photos, virtual 3D models, and real tactile objects like mock-ups, shape models, design models, prototypes and the actual products. The design representations examined in this study were mainly 2D pictures, but some experiments were carried out with virtual 3D models as well. Figure 1 outlines different representation types and fidelities.

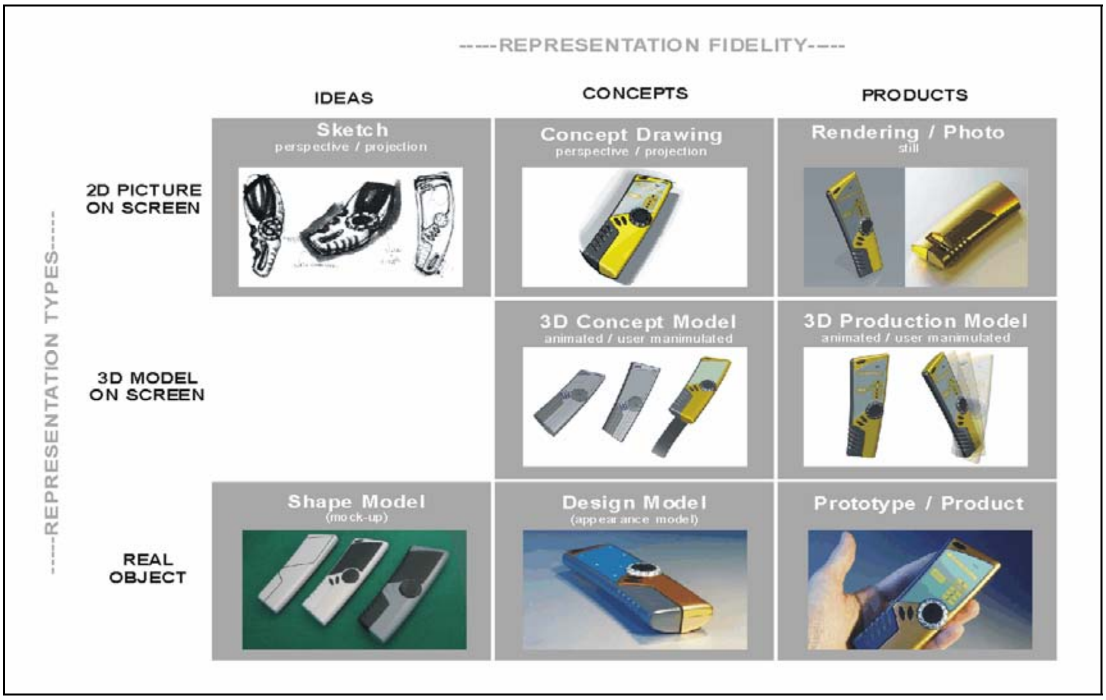


Figure 1: Different representation types and fidelities.

2.2 Visual perception

The human eye provides an elliptic field of vision that extends roughly to 180 degrees horizontally and to 130 degrees vertically. The so-called "useful" visual field extends to some 30 degrees, while the area of high acuity (parafoveal zone) covers only 4 to 5 degrees of the view. The area of the highest acuity in the middle of the view, the foveal zone, covers approximately 2 degrees of the view. Already at 5 degrees the acuity is only 50% of the acuity of the foveal zone and it drops off sharply beyond. Figure 2 illustrates the extent of these areas from the typical viewing distance of the computer screen.

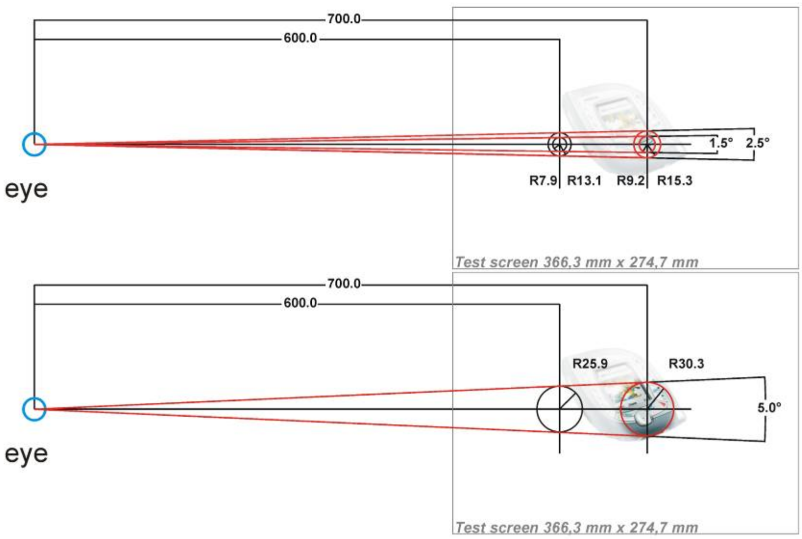


Figure 2: Foveal and parafoveal zones.

These variations in acuity result from distribution of two different types of cells, rods and cones, in the eye's retina, and from their diverse function. The rods are very sensitive in the dark, but cannot resolve fine details, whereas the cones have high acuity, but low sensitivity in darkness. Most of the retina is fairly evenly covered with both cell types, although the rods outnumber the cones by almost 20 to 1. The fovea, however, consists of cones alone. The peripheral vision of the human eye is good for perceiving motion, but the foveal and parafoveal zones of the vision are required for perceiving details.

A gaze path consists of fixations, which are the moments when the eye is relatively still and focused on a certain target, and of saccades, which are the rapid ballistic movements between the fixations. A fixation lasts for 200-300 ms on the average, while the saccades usually last less than 100 ms. The durations given here, especially for fixations, are only indicative: the differences between individuals and tasks cause a high variation. The brain receives information from the eyes only during the fixations, while it is either insensitive to high visual velocities or actively suppresses the vision during the saccades. This is why the eyes are almost constantly moving. In general, more than 150,000 eye movements occur each day for one person (Abrams, 1992).

2.3 Recording eye movements

Observations of eye movements can provide insight on human perception of objects in the surroundings, as there is a close connection between vision and cognition. Gaze positions can be recorded using eye-trackers. In the past, eye-trackers have used some intrusive techniques to follow eye movements, such as contact lenses or electrodes. Current remote eye-tracking techniques are vision based—one or more cameras are used to capture an image of the eye. Gaze position is calculated from the relative positions of the pupil centre and the corneal reflection produced by infrared light shone to the eye (Morimoto and Mimica, 2005).

Gaze path analysis is done with software that can replay the gaze path, show different kind of visualizations and provide statistical data. Figure 3 presents an example gaze path drawn on top of the image of a hedge clipper. Circles represent positions of fixations and lines represent saccades. The radius of a circle illustrates the duration of the fixation. Another visualization method for gaze data are heat maps that highlight the areas of visual interest. Figure 4 presents a heat map on a mobile phone while it was viewed by a participant.

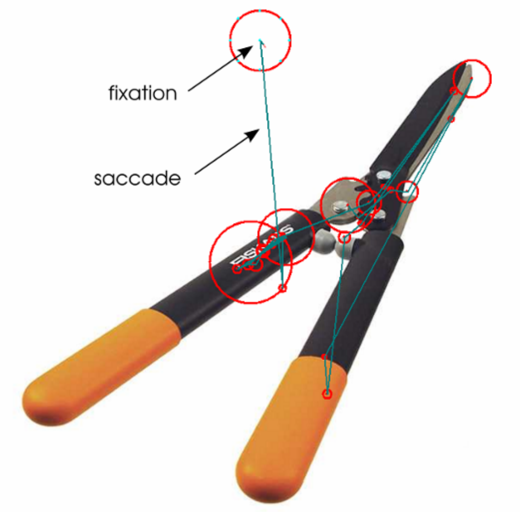


Figure 3: A gaze path visualization.

Figure 4: A heat map visualization.

2.4 Eye movement research

Eye movement studies have started a century ago. They have produced useful information for many application areas, such as usability research, user interface design, and human cognition. Eye movement research flourished with the improved technologies for eye-tracking in the 1970s. The work during the late 1970s was mostly done in psychology and physiology. It focused on exploring how the human eye operated and what it could reveal about perceptual and cognitive processes. When performing everyday tasks, the point of gaze is often shifted toward task-relevant targets even when high spatial resolution from the fovea is not required. Monitoring these eye movements that are made without conscious intervention thus provides us a window into cognition (Liversedge and Findlay, 2000; Pelz et al., 2000).

Eye-tracking technology is being used in an increasing number of applications and research fields (Duchowski, 2003), such as in the studies on perception of physical objects. There is a distinct connection between eye movements and hand movements. Researchers from the University of Rochester measured patterns of eye-hand coordination while manipulating objects like simple blocks (Pelz et al., 2001), and making a sandwich (Hayhoe et al., 2003). While making a sandwich, Hayhoe and others focused on the temporal dependencies of natural behaviour. Perception can be seen as an active process of interaction between an internal schema and the information in the world. People use gaze in a proactive manner: we look at things before we act on them (Land and Furneaux, 1997). Furthermore, people focus at different aspects of an object, depending on the task at hand (Hayhoe et al., 2003) and previous experience (Lu et al., 2001).

Land and Hayhoe (2001) investigated similar natural tasks by examining the relations of eye and hand movements in extended food preparation tasks, tea-making, and making peanut butter and jelly sandwiches. According to their study, gaze usually reached the next object in the sequence of work before the sign of manipulative action occurred. The results indicated that eye movements were planned into the motor pattern and led each action. However, their findings showed that in general the eyes provided information on an “as needed” basis.

2.5 Eye-tracking in image perception studies

A series of eye movement experiments started by Buswell (1935, cited in Babcock et al., 2002) has focused on the perceptual and cognitive significance of eye movements relating to photographs, line drawings, and artwork. While these experiments have demonstrated that observers tend to focus their attention to similar regions in an image, the kinds of eye movements that occur before and during image capture have not been studied. After eye-tracking more than 200 participants while viewing 55 photographs, Buswell found that two types of eye movement behaviour can be observed. In some cases, participants made a succession of brief pauses distributed over the main features of the photographs. Viewing sequences were characterized by a general survey of the image. In other cases, participants made long fixations over smaller sub-regions of the image.

According to his study, people were inclined to make global fixations early with shorter duration, and as the viewing time increased, the duration of fixations became longer with shorter saccades. It was also found that participants often fixated on the same spatial locations in an image, but not exactly in the same temporal order. These consistencies indicated that people tented to focus on foreground elements rather than background elements, and hence did not randomly explore pictures. Buswell also concluded that instructions before viewing objects significantly influenced the perception.

Brandt (1945) investigated the role of eye movements in learning strategies and in the perception of art and aesthetics by analyzing eye movement patterns while looking at advertisements. Both Buswell and Brandt found that there were individual differences in eye movements.

DeCarlo and Santella (2002) studied a computational approach to highlighting the meaningful visual structure in an image. In their study the information from eye movements was exploited to enhance a photograph with a super imposed line-drawing. The user briefly looked at the image, and an abstraction of the image was generated based on the gaze behaviour, combined with automatic edge detection. The elements getting more focus were highlighted and drawn in more detail, while the elements with relatively less focus were rendered with less detail.

By studying the ways the human eye examines complex objects and the principles governing this process, Yarbus (1967) found that eye movements were not simple reflexes tied to the physical features of an image. His studies suggested that the human eye fixates mainly on certain elements of objects that may contain useful and essential information for perception. Elements on which the eye does not fixate do not contain such information (Yarbus, 1967, p. 175). In his well known example, Yarbus recorded the eye movements of participants while they examined I.E. Repin’s “An Unexpected Visitor”. During free viewing, eye movement patterns across seven participants revealed similar areas of interest. Furthermore, he studied how motivation changed the attention. Different instructions, such as estimating the material circumstances of the family, giving the age of the people, and remembering the clothes worn by the people, substantially changed the eye movement patterns for the participants while viewing the painting. In general, the most informative regions were likely to receive more fixations.

Viewing strategies can be affected by several reasons. Henderson and Hollingworth (1998) pointed out that experimental parameters such as image size, viewing time, and image content can cause difficulties in comparing eye movement results over different conditions.

While many studies (such as those by Buswell, Brandt, and Yarbus) had found that participants generally fixate or direct their attention to the same regions while viewing an image, several researchers set out to explore how the semantic features in a scene influence eye movement behaviour (Mackworth and Morandi, 1967; Antes, 1974; Loftus and Mackworth, 1978; Henderson et al., 1999). Noton and Stark (1971a, 1971b) analyzed the chronological order of fixations in an attempt to identify recurring sequences of saccades, called scan paths. In the study conducted by Antes (1974), participants viewed two color photographs, a mask and a coastline. Other than that, in most of these experiments participants viewed black and white line drawings or monochrome shaded drawings of realistic scenes. Again, the general conclusion obtained was that eye movements were not random, and various fixations of participants did land on informative regions in the picture. Furthermore, variability among the participants was also observed, although individuals often followed the same scan paths in specific regions of the image.

In another eye movement study, Molnar (1981) analyzed fixations to find the effect of aesthetic judgments in viewing pictures. Half of the participants in his study were instructed to view the pictures carefully, as they would later be questioned about what they saw. These individuals were designated as the semantic group. The other half of the participants were told that they would be asked about the aesthetic qualities of the pictures. The latter group was called the aesthetic group. Measures of fixation duration indicated that the aesthetic group made longer fixations than the semantic group. However, there was little difference in the magnitude of saccades between the two groups. The longer fixations for the aesthetic group provided an argument that more time was needed to make aesthetic judgments about the pictures, although aesthetic judgments did not influence the angular distance between fixations.

Rayner (1998) studied eye movements while reading texts and processing the information. The study summarized eye movement characteristics during reading. The movements were found to be consistent enough to provide a basis for the development of a formal model of eye movements during reading. When studying eye movements while viewing images, Rayner found that people got the general abstract idea of the image during the first few fixations. The rest of the fixations served to collect the details about the image. Rayner's findings were consistent with those of others stating that eye movement analysis could provide important conclusions about temporal aspects of image perception.

If the parts of a product which attract the most attention are determined, it can be useful in confirming the visibility of a new technical feature or a very important detail, such as an emergency-off-switch on a machine. An already established application of eye movement analysis in car and aircraft cockpit designs or other kinds of instrument panels is based on layout instrument theory. Eye movement recordings are used to optimize ergonomics in design.

Gaze data has been utilized previously in the evaluation of design products by Norbert Hammer and Stefan Lengyel in the University of Essen in the early 1990s. The studies concentrated in trying to determine the connections between a product's details and its semantic features (Hammer, 1992a). Gaze data was analyzed by cumulative dwell time on the product details and by the average gaze path (Hammer, 1992b). Test subjects were students from the University of Essen. Design students were excluded, however, because their perception was believed to differ from others due to their education in the field of design. However, this hypothesis was not included in the study, and there was no attempt to prove it.

Hammer's eye-tracking results explain where test subjects are looking in general when they evaluate different criteria of the products. The studies did not explore the differences between people or groups of people, or connections between centres of attention and product preference. The reports also discuss the so-called "holistic gestalt" phenomenon, where gaze remains focused in the middle of a product, possibly on an empty or smooth area, instead of the details. Hammer and Lengyel (1989) find this phenomenon annoying, since it does not communicate anything about the product details. Hammer and Lengyel conjecture that these gazes represent the evaluation of the product as a whole. This hypothesis has not been verifiable using gaze data alone.

2.6 Eye-tracking studies with experts and novices

Several eye-tracking studies have found varieties of gaze paths between different groups of people, such as experts and novices. Miall and Tchalenko (2001) observed that artists perceive works of art differently compared to novices. During chess playing, while observing the best moving position of the objects, experts were faster and more accurate than the novices (Charness et al., 2001). Experts made fewer fixations and saccades of greater amplitude than novices. The same kind of difference between experts and novices has also been observed in gaze path studies dealing with other activities like surgery (Law et al., 2004) where experts performed better than the novices.

An experiment conducted by Nodine et al. (1991) found that the composition of an image influenced the perception among trained and untrained artists while they looked at paintings. The fixation durations of artists were longer, and their eye movement patterns had a tendency to focus on structural relationships between objects and backgrounds. For untrained viewers, fixation durations were shorter, and eye movement patterns focused mainly on foreground or pictorial elements that conveyed the most semantic information.

2.7 Conclusions

The research reviewed above clearly demonstrates that eye movements can be applied in several research fields to get insights into human perception and cognition. However, the information about the gaze direction alone is not enough without the knowledge of the features of the object and the intention of the observer. Thus, there is a need to combine eye-tracking with other methods of investigation. Also, there is a need to develop methods for distinguishing the features, such as color or shape that capture the consumer's attention.

We wanted to study the above questions further, but it was not clear from previous research what the best approach would be. Therefore several different test setups were experimented with during the project. The research questions were defined in more detail with trying out the test arrangements and equipment.

The empirical studies consisted of three test rounds. The first round covered different viewing motivations with a varying complexity of products. The second one focused on product evaluation. In some parts of this test round eye tracking was supplemented with the think aloud method to obtain information that helps in interpreting the gaze data. Free observation of different product presentation fidelities—drawn sketches, photos and 3D models—were also included. In these two test rounds we compared also characteristics of perception for designers and non-designers. The third round applied knowledge from the previous rounds to design evaluation cases for products of the business partners in the project. In the following sections the test rounds are explained in detail.

3 Exploring Possibilities to Observe Perception of Design Products

In the first test round our aim was to explore different ways to evaluate gaze paths while users are viewing design products. We saw in the studies reported in Section 2 that motivation is a key element in visual perception, so we decided to study perception using a variety of motivating tasks. Most of the earlier studies have focused on perception of pictures or paintings with human figures. With so little past research on perception of artifacts, it was natural to take the results obtained with human figures as a starting point: we wanted to see whether the findings could generalize to design products. In particular, we were interested to see how gaze behaves during the first impression of a product.

The influence of the visual complexity of products on perception was also one of topics on the first test round. We wanted to see how simple things are observed and if there are differences compared to more complex and detail rich products. Therefore five different products were chosen for the test (Figure 5).

Figure 5: Product pictures in first test.

3.1 Test setup

The test was divided in five parts. In the first part participants were not given any particular instructions. In this part we recorded gaze behavior during the first impression of a product. The following four parts gave the participants different motivations for perception. These were to memorize the product and to evaluate its aesthetics, usability and durability. In each part the product pictures were shown in different order for 10 seconds at a time.

The main guideline in designing the test screens was to represent the products as naturally as possible and from the same viewing angle from which the products are seen when they are in real use. Products were also shown in natural size, only the hedge clippers had to be downscaled slightly to fit their image on the test monitor screen.

The participants' gaze paths were recorded with a head mounted SR Research EyeLink eye tracker (Figure 6). The frame rate of the tracker is 250 hz and gaze position accuracy is less than 0.5 degrees. The pictures were shown on a monitor using the iComponent eye-tracking analysis tool that has been developed in the University of Tampere. The tests were video recorded for later review and discussions. The subjects' background information and earlier experiences with the stimuli products were gathered with questionnaires.

Figure 6: Test setup with the EyeLink eye-tracker.

The test involved 20 participants: ten designers and ten non-designers. Ten had a designer education and 10 did not. The mean age of the participants was 33 ranging from 26 to 46 years.

3.2 Analysis and results

For all screens we computed basic gaze parameters, like average fixation durations, fixation counts, saccade counts and saccade lengths. We also analyzed whether these parameters differ between designers and non-designers, between different motivations for viewing, and between products of varying complexity.

3.2.1 Visual complexity and different motivations

The average fixation count for five products in different tasks was used to study the effects of visual complexity. According to a paired-sample t-test, the simplest object (a coffee mug) collected significantly fewer fixations than any other object (see Figure 5). Moreover, the fixation count for the hedge clippers was significantly smaller than that for the camera. These results suggest that a simple product could get fewer fixations (with higher duration) than a complex one. On the other hand, fixation counts for the other combinations of products did not produce any significant difference. Thus, the effect of product complexity needs to be confirmed with a more controlled study.

The analysis of fixation durations indicates that people look at the same product differently depending on their motivation. The task affects their gaze path as well. These findings are consistent with the pioneering study by Yarbus (1967).

3.2.2 Comparing designers and non-designers

We found some differences between designers and non-designers, especially for the memorizing task. In this task, participants were instructed to memorize the products displayed for 10 seconds and be prepared to answer a question regarding the characteristics of the product. Statistical analysis of gaze data showed that the fixation count was significantly higher for designers than non-designers (Figure 7). Consequently, mean fixation duration for designers was shorter than for non-designers. Designers might tend to obtain as much information as possible in shorter time with higher fixation count to memorize objects.

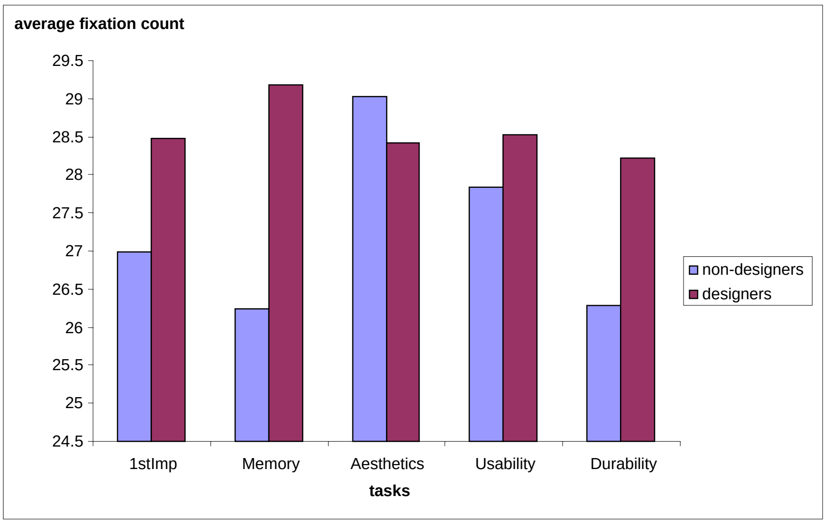


Figure 7: Fixation count for designers and non-designers in different tasks.

Among all the tasks, only the memorizing task produced a significant difference between designers and non-designers.

3.2.3 Different perception strategies

We also analyzed gaze paths in a more visual way. For each product and each participant, the first five fixations were drawn individually as heat maps. These visualizations were then organized in three groups based on how the fixations cover the products. Each group corresponds to a different strategy for starting the perception process. We called these strategies the narrow, holistic, and combined strategy.

People with a narrow strategy (Figure 8 and Figure 9) explore the products with shorter saccades and with fixations that focus more on the same area. People who apply a holistic strategy try to view products so that they get an all-inclusive view of the product as quickly as possible (Figure 8 and Figure 9). Their saccades are longer and the fixations are more wide-spread. The combined strategy seems to mix these other two. Individuals also seem to use a different strategy for different products.

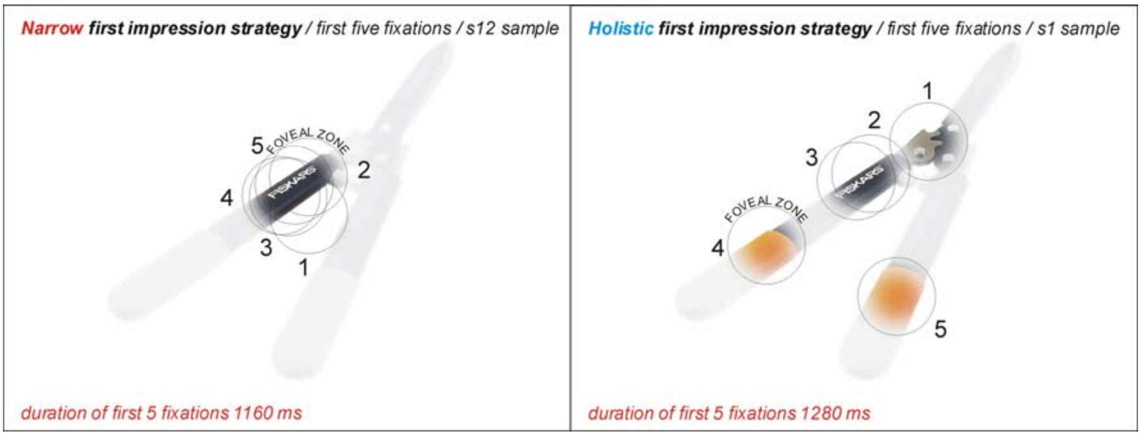


Figure 8: Narrow and holistic first impression strategies for hedge clippers.

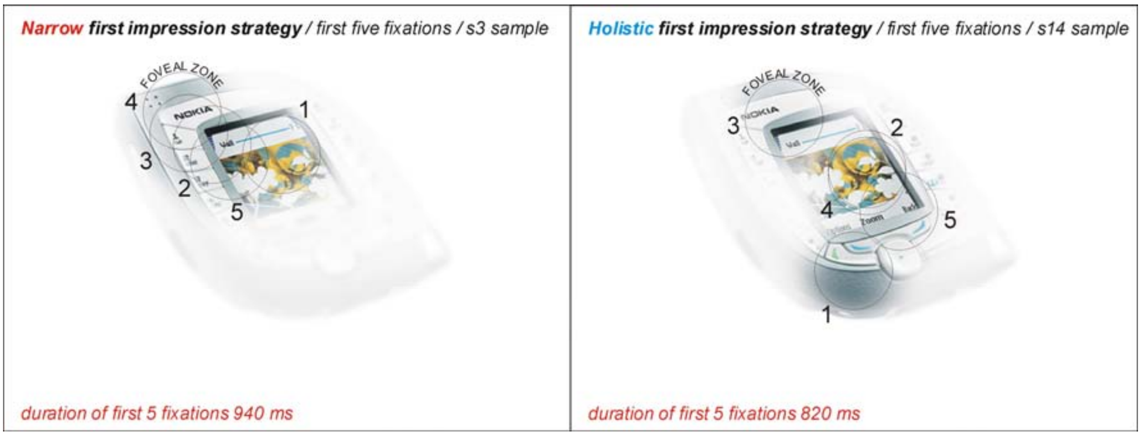


Figure 9: Narrow and holistic first impression strategies for mobile phone.

3.3 Conclusions

The analysis of fixation durations indicates that people look at the same product differently depending on their motivation. Differences were also observed in the gaze paths. These findings are consistent with the pioneering study by Yarbus (1967). The visual complexity also affects the perception but the differences are relatively small. Between designers and non-designers only the memorizing task showed a significant difference between designers and non-designers.

Based on our preliminary results from the visual analysis we suggest that people have at least two different first impression perception strategies: a narrow and a holistic perception strategy. The difference is in how widely the product is covered during the first fixations. The narrow strategy focuses on some details whereas the holistic strategy seems to cover whole product as quickly as possible. We also found a third strategy, called a combined strategy, but it is not clear whether this is really a distinct strategy or an inconsistent use of the two main strategies. These results are still quite preliminary and need to be confirmed with future studies. Similar strategies were, however, also suggested in an earlier study (van Zoest and Donk, 2004).

The first tests gave us valuable experiences about how to succeed in the arrangements of visual perception tests. An important conclusion was that a broad, loosely controlled test setup produced data where it was difficult to find significant differences. In the next test rounds the test conditions were therefore controlled more carefully.

4 Focus on Product Presentation Fidelities and Product Evaluation

In the second eye-tracking test round the research questions were narrowed down to three topics: free observation with different product presentation fidelities, product evaluation, and product evaluation with thinking aloud.

With free observation of different product presentation fidelities we wanted to study if there are perception differences between a drawn sketch and a photograph of a final product. Virtual 3D models of products were also evaluated with free observation. The aim of the test of virtual models was to find a suitable 3D presentation technology that allows gaze path recording with freely manipulated objects. We also wanted to get initial ideas about how the use of 3D models affects the test setup and the analysis of gaze paths.

In the product evaluation part we studied if the gaze would reveal anything about people’s attitudes towards the products. We also studied how the appearance and apparent usability of the products correlated with users’ choices of a preferred product. The gaze data of designers and non-designers was compared to see if they differed in this respect.

The main aim in the product evaluation with thinking aloud part was to study how participants compare and evaluate a pair of product pictures and how they verbalize their attitudes and perceived product attributes.

4.1 Test setup

A new eye tracker, Tobii 1750, was taken into use for the second tests. Tobii 1750 is a remote eye tracker, where the tracking device is integrated into a 17" TFT monitor (Figure 10). This allows less intrusive testing without any disruptive equipment attached to a participant. The frame rate for Tobii 1750 is 50 Hz and the accuracy is 0.5 degrees. Product pictures were shown on the monitor with screen resolution 1280x1024.

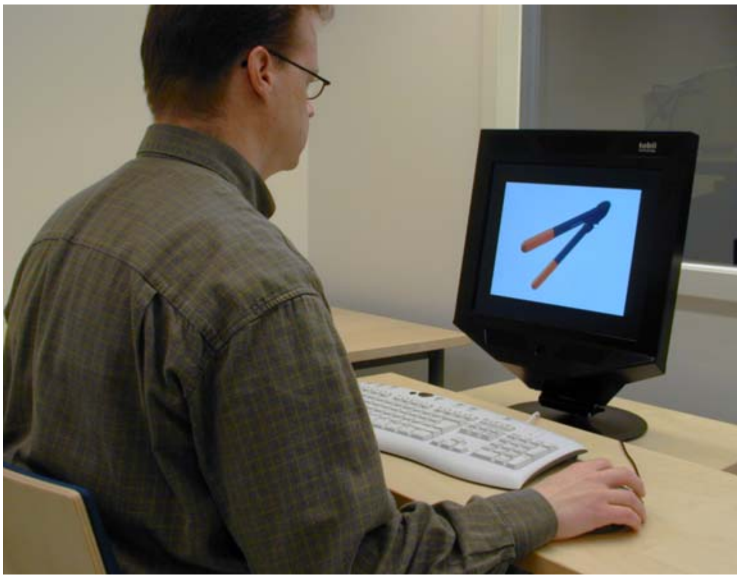


Figure 10: Tobii, a remote eye tracker.

The test image sequence started with the free observation task of drawn sketches and photos of products (Figure 11). The products were two mobile phones, Nokia 6820 and Nokia N-Gage, a Fiskars gardening hoe and a Fiskars axe, in this order. They were shown one at the time for 6 seconds without giving any motivation. Sketches of all four products were shown first, followed by the photographs in the same order.

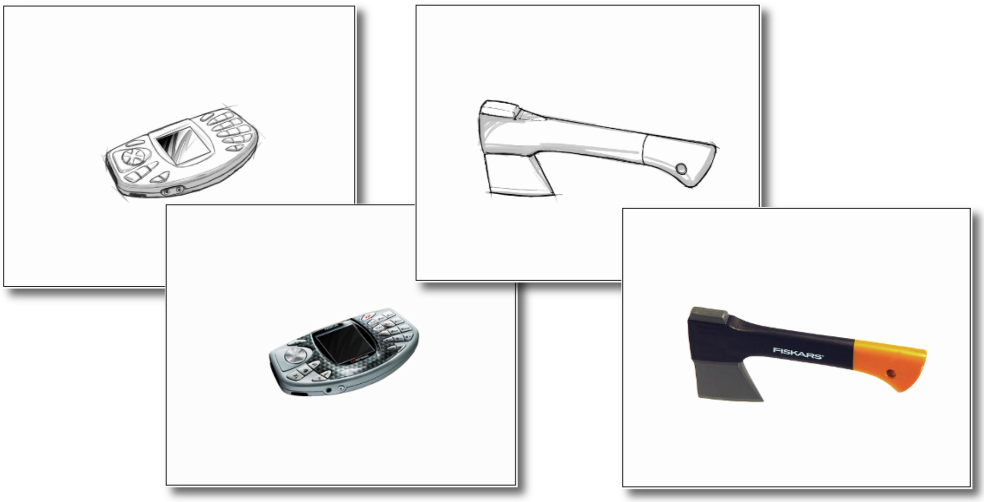


Figure 11: Sample screens from sketches and photos of different products.

For the product evaluation test we selected five pictures of mobile phones that were not marketed in Finland. The brand names were masked. The pictures were then arranged in pairs so that each phone was paired once with all the other phones. The complete set of 10 screens presented each phone four times and equally often on the left and right side of the screen. For each pair, the participant was asked to distribute 10 points between the products, so that the scores indicated how much the participant preferred one product over the over. Each comparison screen was shown for 8 seconds, after which a cartoon balloon with two question marks appeared below the phones to remind the participants to announce the points for each phone. In the end all five phones were shown together and people were asked to pick their favorite (Figure 12).

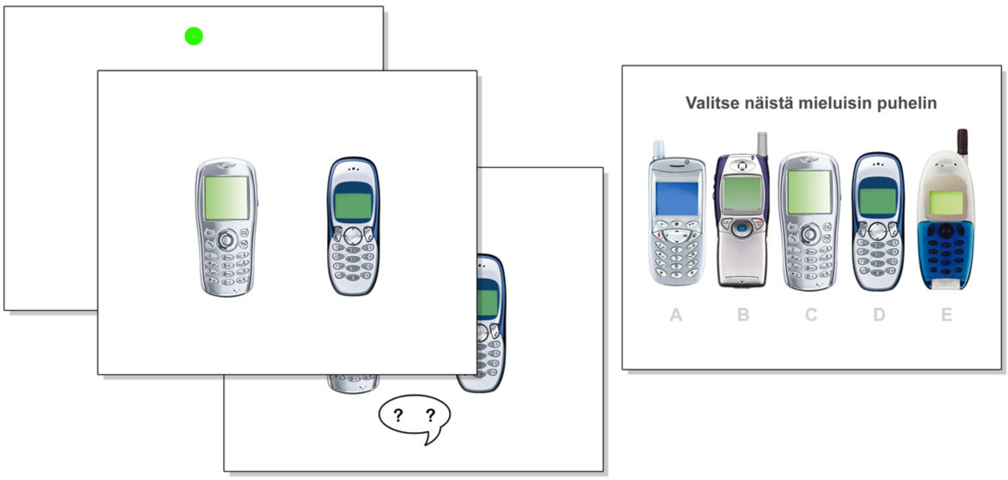


Figure 12: Sample screens from phone evaluation.

In product evaluation with thinking aloud we used a different set of four phones (Figure 13). The test subjects were first asked to look at a pair of mobile phone pictures, and then they were asked to view the phones again with the task "Which one do you prefer and why? Think aloud!" These two tasks were repeated with all the picture pairs. Finally all four phones were shown together on one screen and the subjects were asked to evaluate them aloud on the scale 1–10. Their gaze paths and think aloud argumentations were recorded. Test screen sequences and times were as follows: for the "View as you wish" screens 10 seconds, for the "Which one do you prefer and why?" screens self-determined and for comparison screens self-determined.

The main reason for using verbal protocol analysis in our eye-tracking tests was to understand why participants are looking at certain areas of interest in different tasks. Can gaze path and verbal protocol together elicit the product evaluation process of people? How do people verbalize their attitude and what do they actually evaluate with their gaze?

We followed Ericsson and Simon's principles (Ericsson and Simon, 1984, 1998; Ericsson, 2002) for collecting verbal reports in our product evaluation tests, with the exception that we allowed minimal interaction between the participant and the test operator during the tasks. The test operator acted as a listener who reacted to the participant's speech with minimal responses like "mmh". We proceeded this way so that the participant would feel more natural than "talking alone in a room" with a passive listener. We also considered the earlier experiences and attitudes that the participants verbalized during the test as valid data, because the participants expressed them without any external influence.

Figure 13: Sample screens from verbal product evaluation.

In the 3D test two virtual models of Nokia mobile phones were used (Figure 14). The models were in Cycore Cult3D format, which is mostly used for presenting 3D material in WWW pages. The virtual models in the test were presented with a Cult3D viewer from Microsoft PowerPoint. Participants rotated the models freely with a mouse without any motivating task.

Figure 14: The 3D virtual models used in the test.

After all three parts of the test were completed, each participant filled in a questionnaire about their background including education, profession, age, gender, and possible problems with eyesight. For each phone in the product evaluation part the subjects were asked to mark on the questionnaire positive and negative details in each phone and to give scores from 1 to 10 to the appearance and apparent usability of each phone. Similar scores were given to the personal importance of five possible factors affecting the choice of a mobile phone: brand, appearance, price, usability and technical features. Finally, the participant's familiarity with all the products was inquired.

There were 32 participants in the test. For free observation, gaze data was obtained from 26 participants (13 designers, 13 non-designers), for product evaluation from 28 participants (14 designers, 14 non-designers) and for product evaluation with thinking aloud 24 participants (12 designers, 12 non-designers).

4.2 Analysis and results

4.2.1 Differences on drawn sketches and photos

Statistical fixation analysis revealed no systematic differences between drawn sketches and photos. According to paired-samples t-test, significant differences in fixation count and average fixation duration exist between a sketch and a photo only for Nokia N-Gage, but not for any of the other presented products (Figure 15).

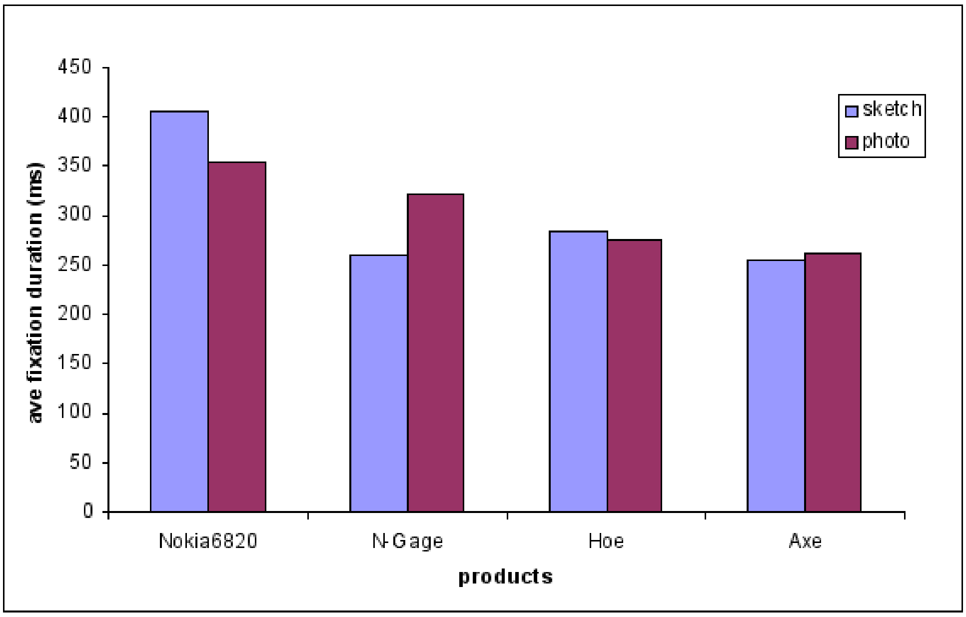


Figure 15: Average fixation durations on sketches and photos.

We also analyzed whether the various areas in the product images received a similar amount of attention in the sketches and photos. Some differences were observed, but they were anticipated, since (1) the sketches had fewer details, and (2) the sketches also had fewer areas of interest – for instance, brand names were omitted from the sketches. Thus, this analysis did not produce interesting observations.

4.2.2 Product evaluation

The results from analyzing the first fixations in each comparison screen suggest that the western reading direction may play a role in viewing and comparing the pictures in a left-right layout. The left picture was looked at before the right picture in 78% of the screens. Seven people out of the 28 participants always looked first at the left picture. Becoming familiar with the pictures seemed to decrease the dominance of the left side towards the end of the 10 screens, though. The left picture also had a higher fixation count average than the right picture and the average total viewing time spent on the left phone was also higher than on the right one. Although the way of viewing seemed to favor the left phones, overall the left and right sides were rated equally.

Our main goal in this part of the study was to find out whether gaze data (fixation counts or fixations durations) could serve as an indication of product preference. While a significant correlation was found for some of the five phones, this was not the case for others, and in the extreme case there was negative correlation. Thus, based on our tests, gaze data alone cannot be used to make conclusions about product preference.

4.2.3 Product evaluation with thinking aloud

The verbal protocol was transcribed and transcriptions were organized by segments of (verbal) idea units or intonation units. An idea unit is the mental unit that is the object of investigation. (Holsanova, 2001) An intonation unit is a sequence of words combined under single, coherent intonation contour, usually preceded by a pause. (Chafe, 1980, 1987, 1996).

On the average, non-designers used more time to evaluate the product pairs than designers.

Usually participants started their visual observation during the "View as you wish!" task with a quick exploration of both phones. They seemed to first either look into the space between the phones, or to view both phones with a few fixations to identify what they are evaluating.

In the later parts of the test some subjects started to examine one of the phones with a long sequence of successive fixations. The verbal idea units synchronized with these long glances reveal that subjects examined and evaluated a single phone and its attributes. The attention was focused on one phone without comparison. Then comparative evaluation was continued.

Figure 16 illustrates an example of successive fixation units and verbal idea units. They are parts of the following verbal sequence where the participant wonders where power-buttons are located and how easy they are to use.

"… then ( )
I can't right now see where that power button is and how easy it is to use ( )
if it is there associated with C-button in that it is at least not ( )
in that rounder model it is not at least in C-button which is a good thing ( )
ehm ( ) … "

In Figure 16, the sequence is divided in six frames. Each frame contains the fixation numbers in the top left corner; red fixation numbers indicate that the subject speaks during those fixations. The fixations of each frame are visualized as heat maps.

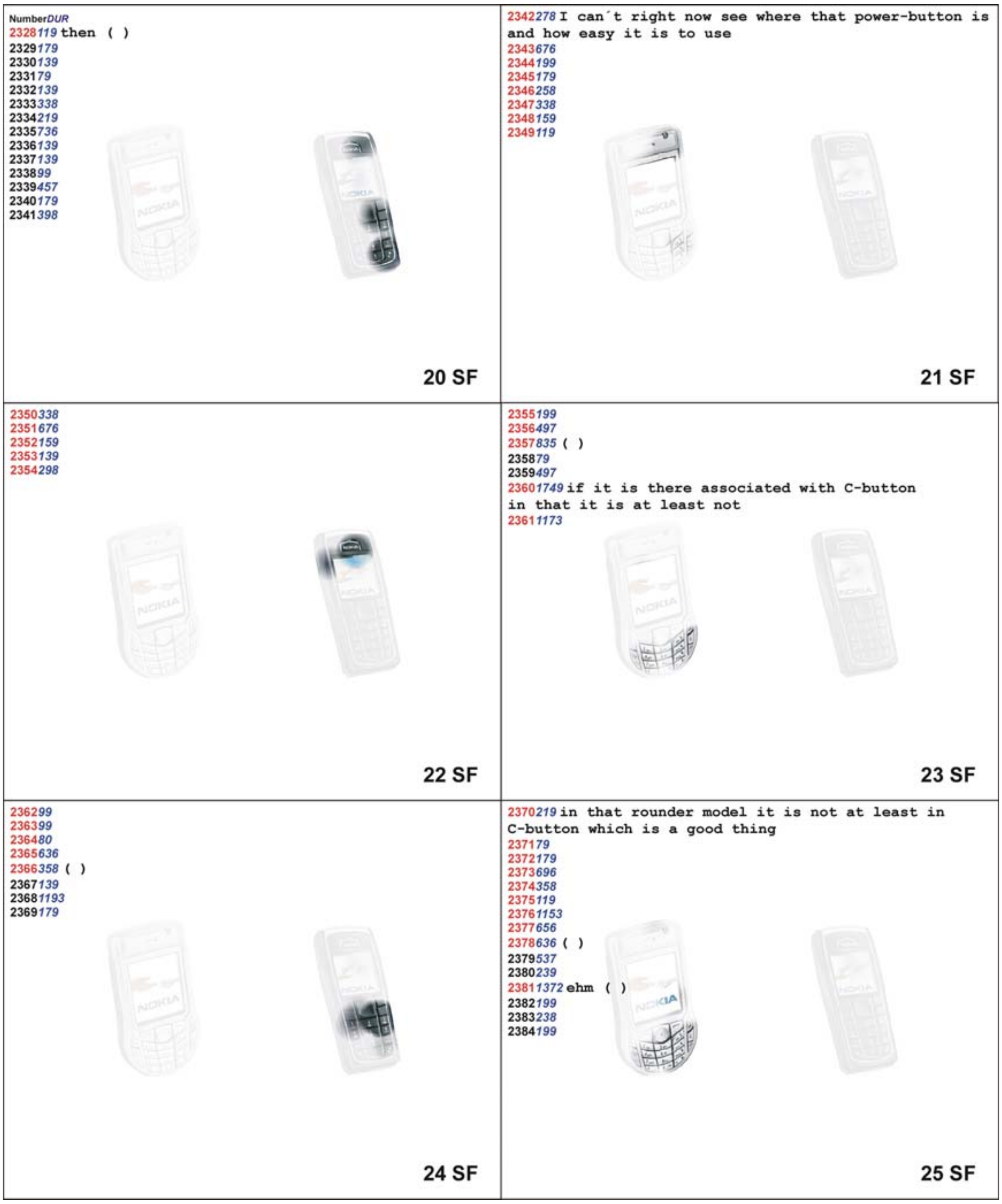


Figure 16: 20–25 successive fixations units / verbal idea units of one participant during "Which one do you prefer and Why?"

These phases of exploring and evaluating can be called identification, exploration and focused examination phases. During these phases the visual attention switched from general to specific in a reiterative way. This result is in line with the studies of Yarbus (1967) and Holsanova (2001).

4.2.4 Differences between designers and non-designers

In the free observation task, where only one product was shown at the time, no significant differences in the gaze data (fixation counts, fixation durations or saccade lengths) were found between designers and non-designers.

In the product evaluation task there were statistically significant differences in fixation counts between the groups (Figure 17). Designers made fewer fixations on four phones. This is in line with earlier studies (Charness et al., 2001), where it has been found that experts are able to evaluate a situation with fewer fixations. In our study there were no differences in fixation durations or saccade lengths.

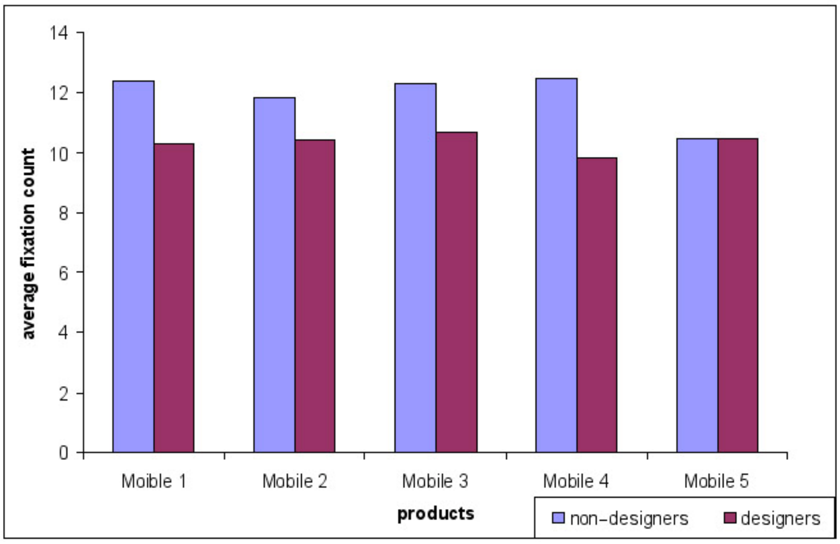


Figure 17: Fixation counts in product evaluation task.

Some more differences were found between designers and non-designers when the products were divided into different areas for the analysis. For example, the “back blade” area of the axe got a significantly higher number of fixations, with a longer average duration, by the designers than by the non-designers. The fixation count was also significantly higher for the orange handle of the axe and for the number pad of the Nokia 6820 by the designers than by the non-designers.

In the product evaluation task the designers’ average fixation count correlated well (r = 0.90) with the appearance evaluation scores in the questionnaire, but not at all with the evaluation of apparent usability (r = –0.57). For the non-designers appearance (r = 0.70) correlated worse than apparent usability (r = 0.84). This might suggest differences in evaluation interests and motivations.

The average fixation durations did not correlate with the ratings as well as the average fixation counts. For the designers the correlation was fairly high (r = 0.85), but for the non-designers very low (r = 0.36). Designers may have based their comparison and personal preference much more on appearance than on apparent usability, whereas the non-designers have considered both attributes.

In product evaluation with thinking aloud, the average fixation durations of designers were higher than those of non-designers in all five tasks. The reason for this could be that designers have learned to look differently while practicing drawing. This is a conjecture that requires further study, but some evidence was provided by Miall and Tchalenko (2001, p. 38), who observed that for artists the average fixation duration was higher than for novices during the drawing of a sketch of photographed faces.

4.3 Conclusions

The findings in this study show that it is possible to find correlations between gaze data and different aspects of design evaluation. These findings may prove useful in developing the combined eye-tracking and design evaluation methods further. When applied to real design cases these methods could assist a designer in various stages where the visual aspects of the design need to be evaluated. Hammer and Lengyel (1991) summarized how eye-tracking can help a designer: "It is undoubtedly useful for measuring attention, to determine the most attractive areas of a product." With given tasks, it "can indicate which product elements carry the given meanings or brand identification". However, eye-tracking was not found useful when the meanings were carried by the whole of a product.

One possible future application could involve comparing the intended product claims to the areas of the product that attract visual attention. The results could be utilized to define areas and details requiring special attention in a design facelift. Another facelift approach could investigate which product details carry the cues of brand identity and how well the existing products stand out from the mass of similar products.

These same approaches could also be applied in studying new visualized design concepts. A study of competitors' products or the company's own older products could also provide valuable data for the basis of a new product concept.

Evaluative verbalizations (verbal foci) synchronized with fixations (visual foci) shed some light on why participants were looking at certain areas of interest on the phones. A temporal relationship was observed between the visual and verbal foci in the evaluation: verbal focus followed visual focus with a delay. Participants fixated first, thought (verbally encoded) and verbalized later. However, verbal foci cannot always be connected to visual foci mainly because participants continued gazing during the pauses of their verbalization. Identifying different types of temporal relationships between visual foci and verbal foci is an interesting topic for further examination. The eye movement and think aloud protocol test and analysis method presented here needs to be refined, because it is very laborious and time consuming to analyze.

Virtual 3D model tests revealed that more methodological research and technical development has to be done before practical gaze path analysis for virtual models is possible. Currently the only available analysis method with replay video of the screen with gaze points is tedious and inaccurate. Analysis could be improved with fixation data collection in 3D space. This data could be used for area of interest studies on defined areas of the 3D model.

5 Brand Recognition and Product Preference Study—Case Nokia

This study concentrated on testing the previously refined methods in a real design evaluation case involving Nokia mobile phones. The aim was to examine whether these methods could provide answers to more precise research questions. The questions were selected from a set of proposed case studies by Nokia's representative. The test was performed together with the Fiskars case study (presented in the next section) with the same subjects.

The case study aimed to answer three particular questions:
1. Do the Nokia products stand out from the mass of similar products?
2. Which product details do communicate the Nokia brand and how well do they do it?
3. What is the role of these details in influencing product attitude and product preference?

5.1 Test setup

The test was performed with the Tobii 1750 eye tracker and analyzed with Tobii's ClearView software together with common image processing and statistical software. Nine phone pictures were selected for the test. Three of them were Nokia phones and the other six were of Siemens, Motorola, Sony Ericsson, Nec and Sagem brands. Logos were removed and the colour schemes and sizes were unified.

The test comprised of the following tasks (Figure 18):
- free viewing of all the 9 phones at the same time,
- comparing 3 phones at a time to each other (9 screens),
- selecting a favourite from the 9 phones, and
- identifying a Nokia phone among 4 phones (2 screens).

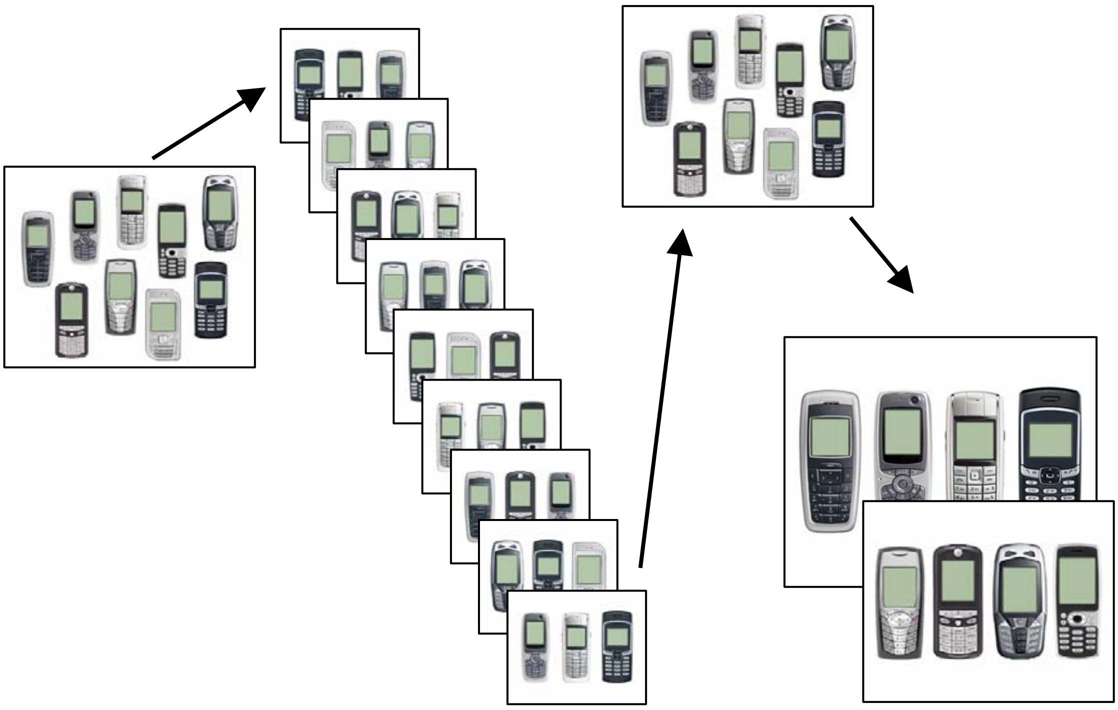


Figure 18: The test sequence: free viewing, comparisons, selection, brand recognitions.

The free viewing screen presented all 9 phones in the beginning. The objective was to record the initial attention on each phone in order to enable comparisons of this attention to product attitude and preference.

The task in the comparison screens was to place three phones at a time in order from the most to the least liked phone. The answers were given aloud according to phone locations. Each phone appeared three times in the whole set, once in all three locations, and one of the three was always a Nokia phone. The process produced a rank order of the products.

After the comparison screens the same initial screen of all 9 phones was shown and the subjects were asked to pick their favourite phone. The goal here was to identify the most salient details and to compare the attention and selections to the data gathered during the initial free viewing task.

In the end there were two brand recognition tasks. The task was to identify a Nokia phone among 4 phones. The first screen in this task included two Nokia phones while the other had none.

The test included both automatically timed and manually changed screens. The timed screens were meant for enabling equal comparisons between some screens and between different people. The manually changed screens were used when the subject was expected to give answers. In some cases a timed screen was followed by the same screen with manual change to obtain both the comparable time and the free answering time.

After the test the participants' background data, possible familiarity with the test phones and recognition of the brands was collected with a post questionnaire.

The test results are based on 15 subjects, 11 females and 4 males, whose ages varied between 20 and 33 years the average being 24.5 years. 11 persons owned a Nokia phone and 4 had a Siemens.

5.2 Analysis

The ClearView software was utilized to produce cumulative gaze duration heat maps and to export the numerical data to statistical software according to predefined areas of interest (AOI). These areas were defined as slightly larger than the actual phones in order not to exclude fixations at the phone contours. ClearView's default settings of fixation size (50 pix) and minimum fixation duration (70 ms) were used in this process.

Two kinds of heat maps were made from the free viewing, phone selection and brand recognition data. One type illustrates the gaze duration and the other shows the percentage of people who have looked at the most eye-catching areas in the screen. These heat maps were compiled in groups according to the given answers and further simplified and clarified by image processing.

The statistical data included all the fixations in order of occurrence, durations in milliseconds and locations according to areas of interest. This data was then used to define the phones viewed first in each screen, and to compute the fixation counts, durations and average durations on each phone.

The answers given during the comparison screens were calculated as a rank for each phone. Another phone ranking was created from the favourite phone selections. Finally, the ranks, the selected phones and the gaze data statistics were compared to each other and their relationships were examined.

5.3 Results

5.3.1 Free viewing and first impression

The phone close to the center (G, see Figure 19) received most attention (most fixations and longest total duration) in the initial free viewing screen. It also had the longest average fixation duration and 7 of the 15 people looked first at this same phone.

One of the Nokia phones (C) received first looks from 6 people and it was also the third most often and longest looked phone. Nokia phone A had longer average gaze duration than the other two Nokia phones (C & L), but it was still only the fourth longest among all phones.

In the initial viewing the Nokia phones did not seem to attract more attention than any other phone (Figure 19). However, it may be that the central location in the layout plays a meaningful role. A control group with a different phone layout would have been needed to verify this.

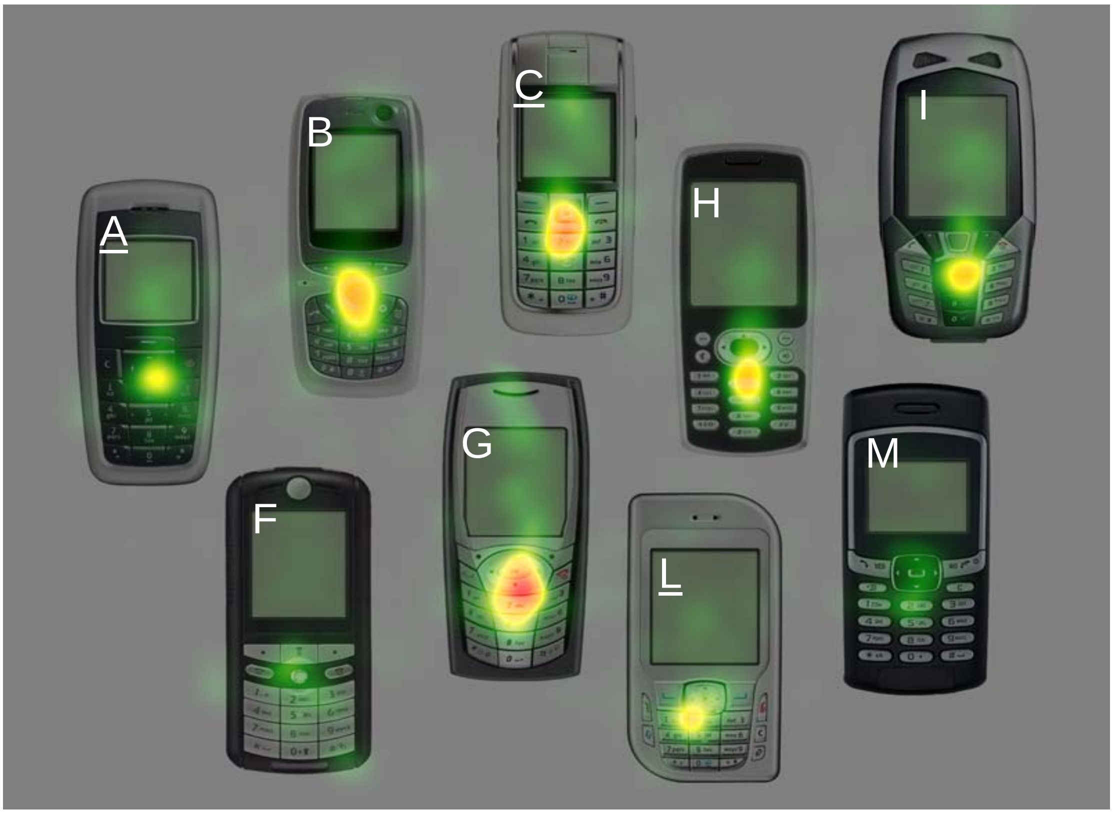


Figure 19: Free viewing (10 sec). The heat map presents the cumulative gaze duration of all subjects. The letters of the Nokia phones are underlined (letters were not part of the test screen).

5.3.2 Brand recognition

All but one person recognized one or the other Nokia phone in the first screen with four phones. In the second screen, which had no Nokia phones, 13 out of the 15 persons elected the same phone (G) as a Nokia model. This same phone was also ranked first in the phone comparisons and selected most often as personal favourite.

The recognition and decision making for the two Nokia phones seemed to occur with different pace. The C-phone, identified as a Nokia phone by 5 people, gathered attention and was chosen faster in general than the A-phone which got 9 votes. The former stood out from the other phones already during the timed 10 second screen while the attention on the latter phone became obvious only in the following manually timed screen (Figure 20).

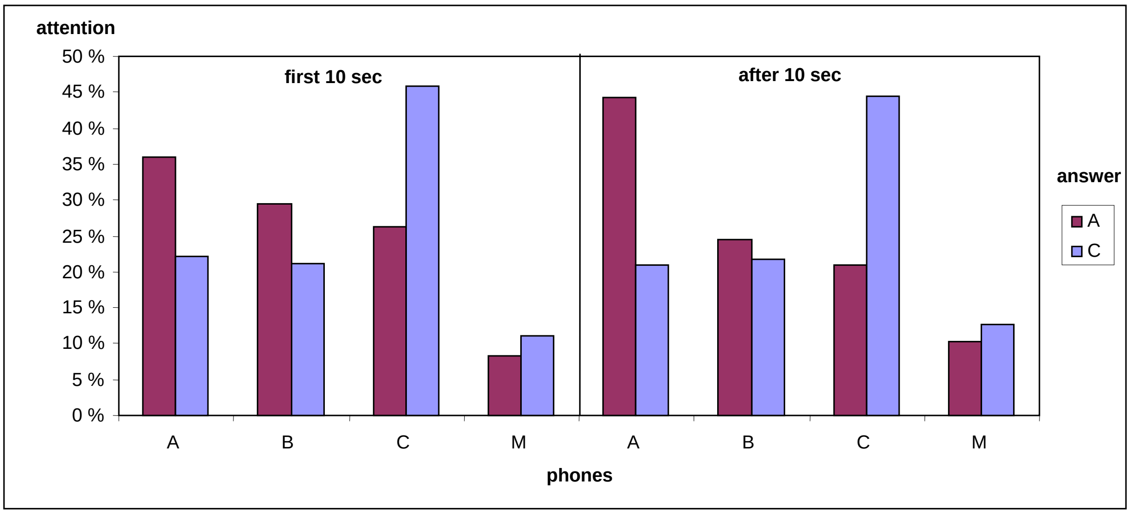


Figure 20: Brand recognition 1 (phones A, B, C, M). Average dwell time percentage on each phone divided according to which phone is regarded as a Nokia phone. The blue C bar stands out from other blues during first 10 sec, whereas similar separation becomes clear for red bars only after 10 sec.

In the post questionnaire all 15 people answered either "yes" or "maybe" in the brand recognition question and proclaimed the G-phone's brand as Nokia. Only one of the actual Nokia phones (L), the one with re-employed recognizable contour, was well recognized with 12 "yes" or "maybe" answers. These answers for the other two Nokia phones amounted to 7 and 5, both with mainly "maybe" answers and only a single "yes" answer.

The heat maps of the brand recognition screens implied that people were trying to find the brand cues mainly from the call control and navigation keys (Figure 21). The role of the whole form could not be determined with this test.

Figure 21: Brand recognition 2 (phones G, F, I, H) – no Nokia phones (10 sec). The heat map presents the percentage of overlapping fixations of those 13 subjects who answered "G . The white edge shows the area viewed by 54% of the subjects.

5.3.3 Attitude and preference

The comparison screens and task were utilized in order to create a rank order for the phones. This order was to be used for verifying the phone selection result. In general the Nokia phones were valued in the comparison task being ranked 2nd (A), 3rd (C) and 4th (L).

The two most popular phones in the selection task turned out to be the same as the two best ranked phones from the comparison task: phones G and A, respectively. The first ranked phone (G) was selected by 6 people and the second ranked phone (A) by 4 people. 5 other phones were picked by a single person and 2 phones by none. Peculiarly nobody preferred the C-phone in the selection task even though the gaze data indicates attraction above the average and the comparison task ranked it as the 3rd most liked phone. Possibly the C-phone ended up being the second best option for several subjects.

In the phone selection screens most of the attention was again directed at the location of the call control and navigation buttons, but in a more narrow fashion (Figure 22) than in the brand recognition task. The empty displays also received some attention along the center line of each phone. Both facts could indicate viewing of the whole phone. However, the average fixation duration was 10% shorter than in the free viewing screen, which could imply a slightly more detailed and specific evaluation than in the more general initial perception.

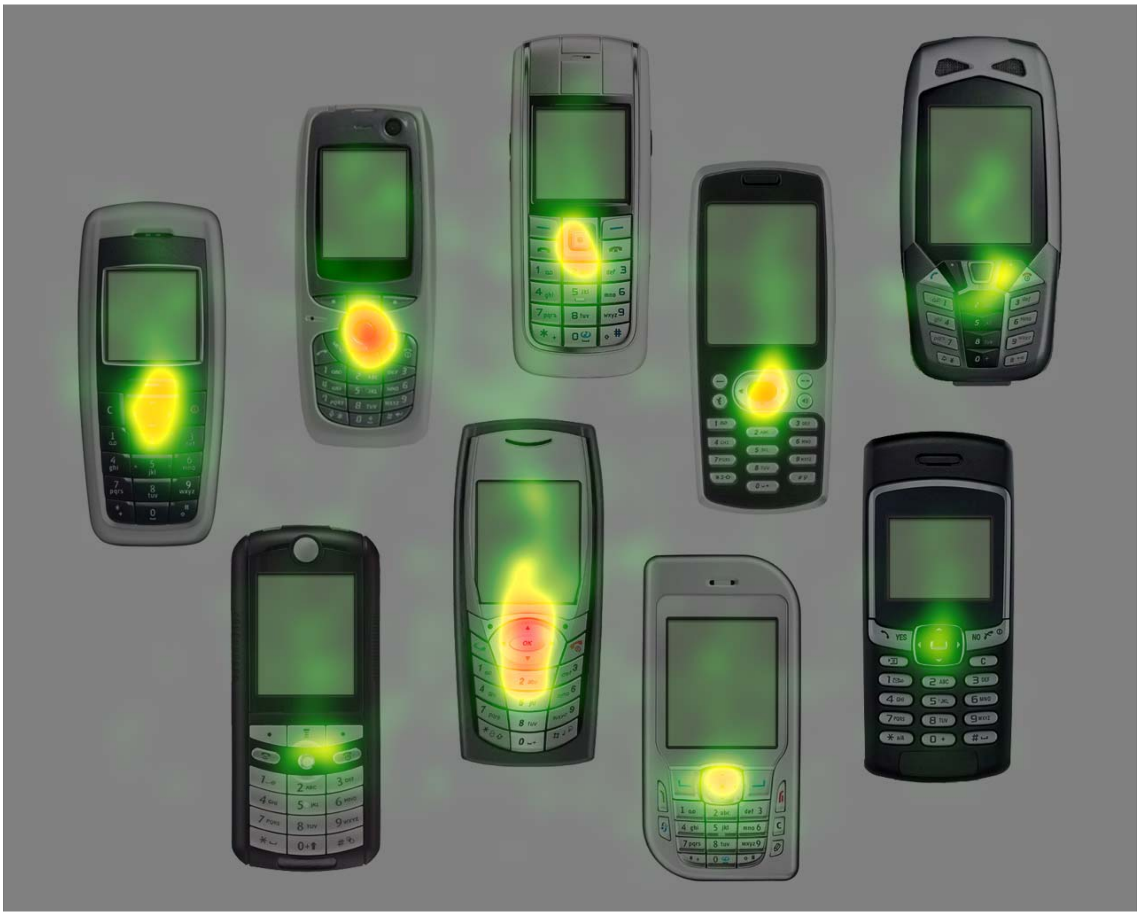


Figure 22: Phone selection (10 sec). The gaze duration of all 15 subjects.

5.3.4 Summary

The actual Nokia phones did not attract the gaze initially particularly more than the other phones. However, the clear eye catcher and favourite phone was mistakenly believed to be a Nokia phone by every subject. The call control and navigation keys were found meaningful in brand recognition, although it could not be determined how well they communicate the brand. The already familiar and recognizable contour seemed to work better in this regard.

5.4 Conclusions

The analysis of the free viewing screen revealed how the various phones caught the initial attention. The comparison screens provided data for ranking the phones and for further analysis of the role of various details. Phone selection data was compared to both of the previous and it revealed the differences and similarities between phone attitudes and final preference as well as some differences in attention build-up. Nevertheless, it has to be taken into account that the screen layout may play some role in these results, especially since the first and third ranked phones were located near the center of the screen.

The brand recognition screens gave some clues about brand related details and about how recognizable some of the phones are. The roles of the whole form and the details could not be determined in this test, but they both seemed to emerge at some stages in different tasks.

This case test proved to be a valuable experiment in examining the usefulness of the method. It provided some answers to most of the case questions and pointed out needs for further development with those that it could not answer.

6 Product Attribute Study—Case Fiskars

The motivation for this test was to understand what subjects look at and how they justify their judgments when evaluating a product photo pair according to a beforehand defined attribute. As discussed in Section 3, knowing the motivation for a viewing task helps in analyzing the gaze data. The given attributes were effectiveness, durability, ergonomics, and aesthetics. They were defined together with representatives of Fiskars. These evaluation attributes are related to some of the product claims of Fiskars.

For this study our research questions were:

- What areas of interest or product details do subjects look at and compare when they are evaluating effectiveness, durability, ergonomics, and aesthetics?
- How do they argue and verbalize their attitude? And what specific characteristics and features do they verbalize when they evaluate and compare product pairs according to different attributes?

6.1 Test setup

Eye movements were recorded with the Tobii 1750 eye tracker and think aloud protocols were captured with a video recorder. Visualizations of gaze data were made with iComponent (v.3.22) software and Photoshop. Numerical gaze data was exported from iComponent and ClearView software and analyzed with Excel.

Subjects practiced first the test procedure and the think aloud method. After practicing the subjects were first asked to view three different product pairs without specific evaluation criteria. The products were hedge shears, loppers, and bypass pruners (shown from left to right in Figure 23). The subjects were then asked “Which one is more effective and why? Think aloud!” Similar tasks with the attributes “durable”, “ergonomic”, and “aesthetic” followed. Finally the product pairs were shown with the question “Which one would you select for yourself and why? Think aloud!” The order of product pairs in the different tasks varied.

For the free viewing screen the time was 10 seconds and for attribute based evaluation screens and preference screen the viewing times were self-determined.

Figure 23: Example of test task sequence (“More effective”).

4 participants out of 17 had to be left out from the analysis because of poor gaze data or mistakes in the test protocol. From the 13 analyzed participants, 10 were female and 3 were males. Their average age was 24.7 years.

6.2 Analysis and results

Analyzing verbal protocols synchronized with gaze data is a slow and tedious process. Careful analysis has been done for bypass pruners. For hedge shears and loppers the analysis is still ongoing.

6.2.1 Areas viewed during evaluation of the bypass pruners

The data was visualized as cumulative heat maps, and the distribution of cumulative fixation duration on areas of interest (listed in Figure 24 and shown in Figure 26) was computed.

As Figure 24 shows, the handles drew most of the attention (always close to 50 % or more). Particularly in the evaluation of ergonomic quality the subjects focused almost entirely on the handles. The relative importance of the mechanism increased in the evaluation of durability. Compared to the handles and mechanism, blades went largely unnoticed, but they had a small role when the subjects selected their personal favorites.

The heat maps in Figure 25 show the distribution of attention (on the average) in the various areas of interest between the two objects in each screen.

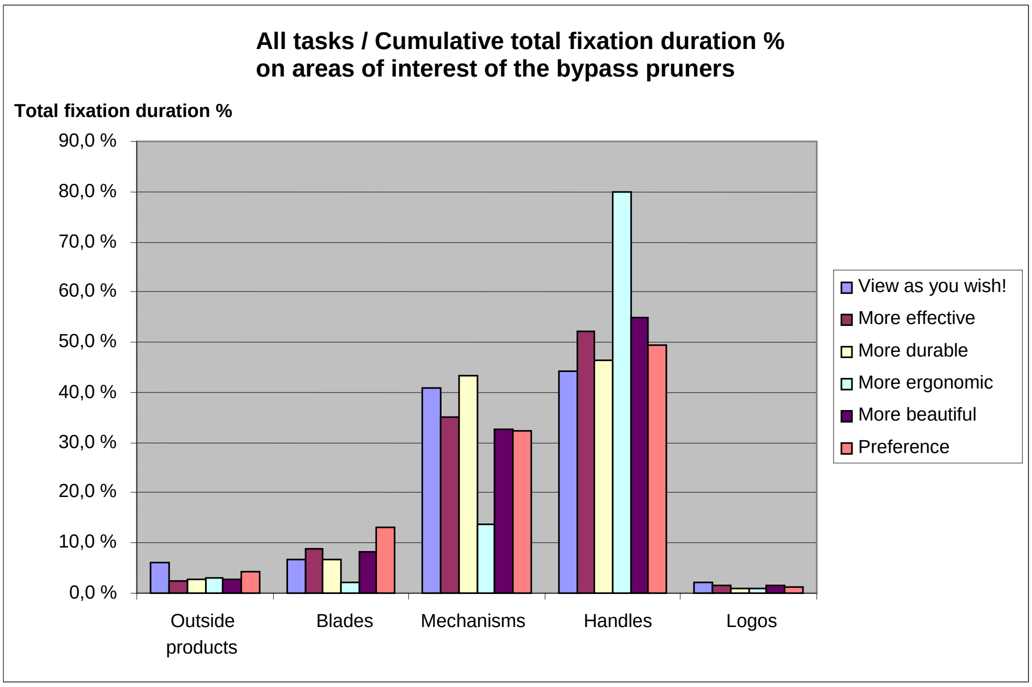


Figure 24: Cumulative average fixation duration percentage on areas of interest of bypass pruners in different tasks.

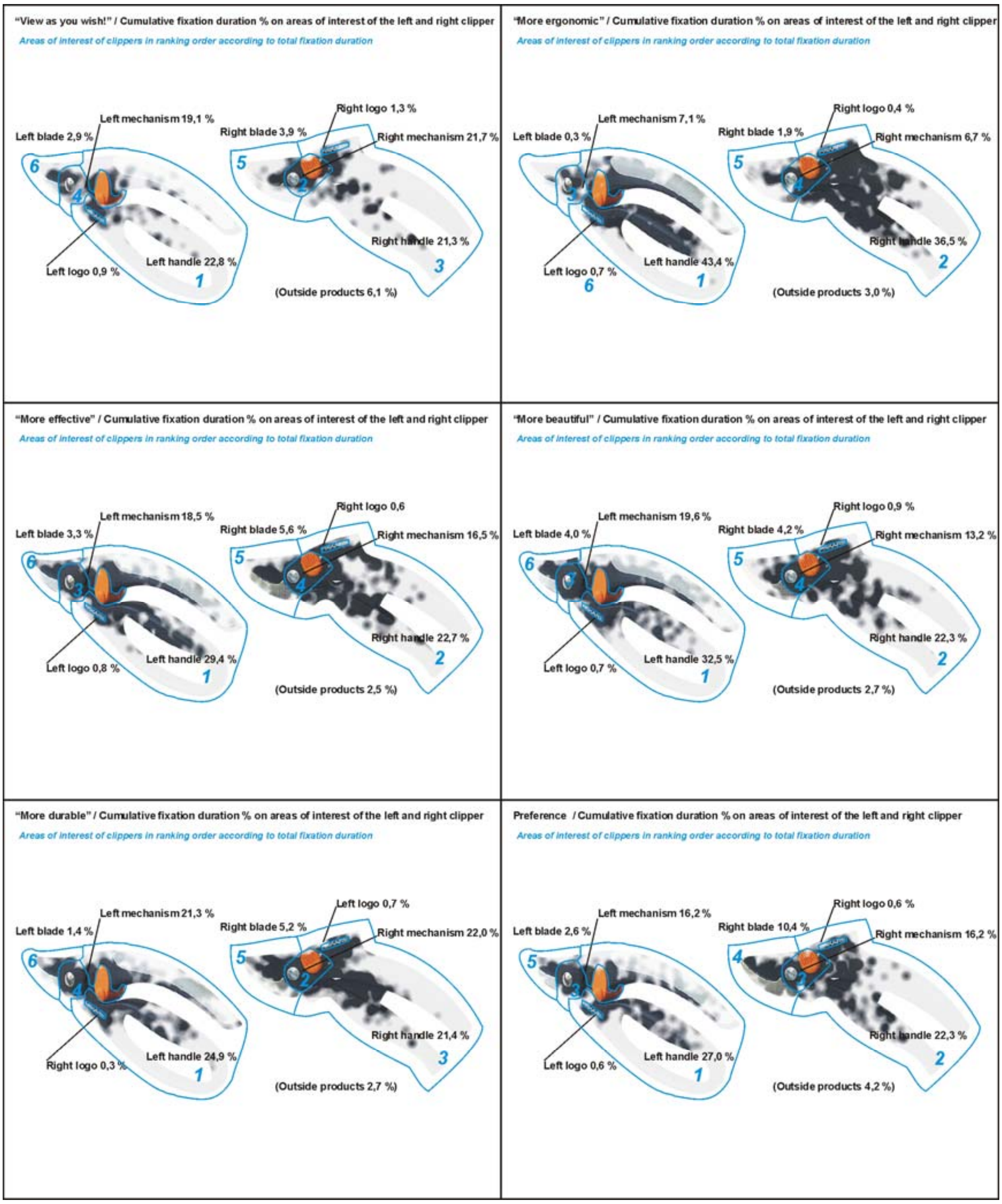


Figure 25: Cumulative heat maps and distribution of cumulative fixation duration percentages on areas of interest of bypass pruners in different tasks (all participants).

6.2.2 Think aloud protocol analysis

Figure 26 tells how participant 2 verbalized his evaluation. The figure also shows how his visual attention was distributed between the areas of interest. Similar visualizations were produced for the different tasks and participants, and they yield mostly qualitative data that helps in interpreting the numerical gaze data.

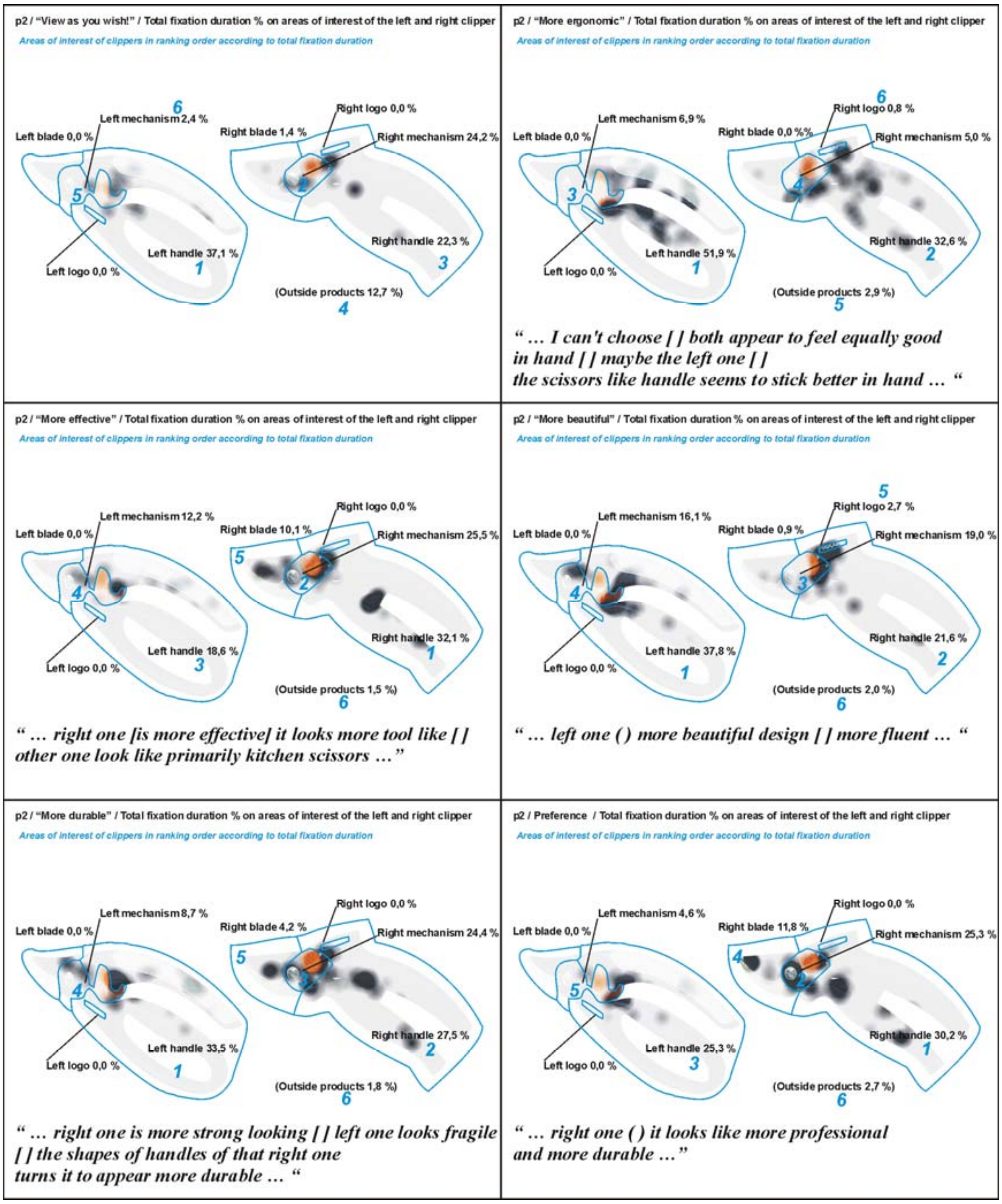


Figure 26: Heat maps and distribution of cumulative total fixation duration percentages on areas of bypass pruners and transcription of verbal protocol (all tasks, one subject).

6.2.3 Fixation durations and preferred pruner

When subjects evaluated the bypass pruners for each of the four product attributes, and also when they chose their own preferred product, they consistently looked for a longer time (longer total fixation duration) at the pruner that they considered better according to the given criterion. Figure 27 shows, on the average, the distribution of fixation durations between the left and right pruner (they were always shown in the same order). The difference was also significant on the level of individual participants and tasks: in 86.2 % of the cases participants looked longer at the pruner that was assessed better. Interestingly, in the free viewing condition no such difference was observed.

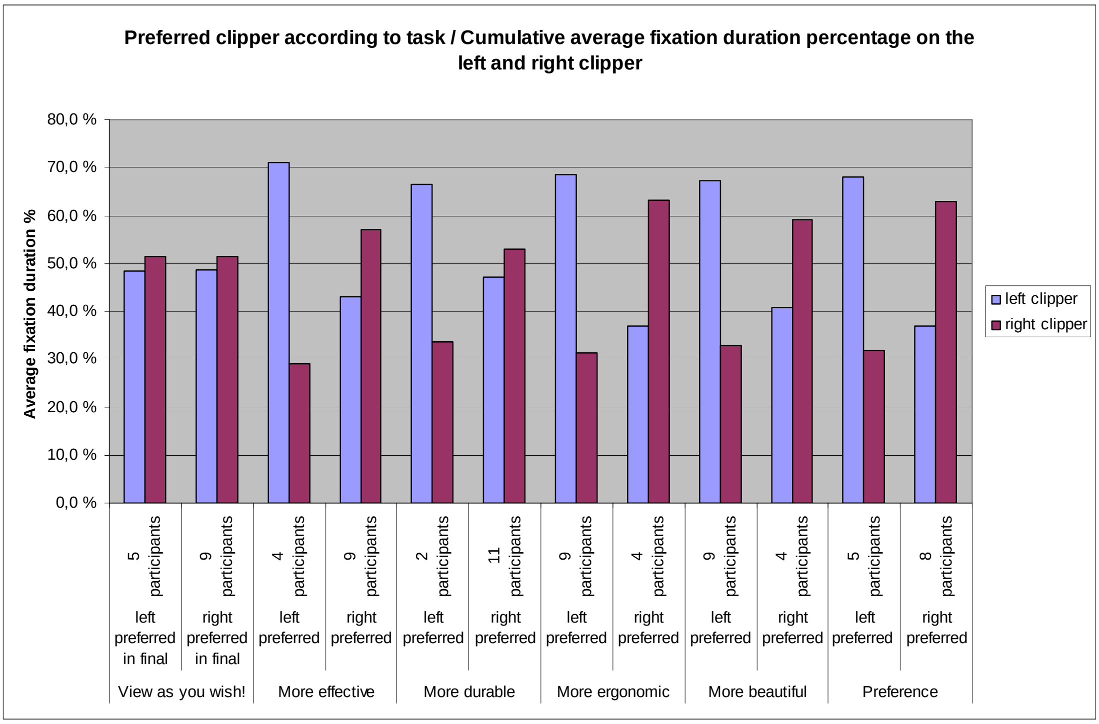


Figure 27: Cumulative average fixation duration percentage on the two pruners according to preference (all tasks).

6.3 Conclusions

The difficulty in drawing more general results from the analysis of individual cases lies in defining areas of interest of two different kinds of products so that the areas defined are comparable in terms of area, contents, function, and visual message. Sometimes these corresponding areas are located in totally different places and sometimes they overlap with other areas.

When participants evaluate one attribute at a time it is possible to detect which areas of interest the participants gaze at. This is an indication to designers wishing to convey the image of, say, an ergonomic product, on what part of the product they should pay attention to.

The results obtained above for bypass pruners need to be verified by conducting the same analysis to the other two product pairs of this experimental test. Preliminary results of the analysis of hedge shears and loppers seem to be in line with the results for bypass pruners.

7 Test Scenarios

Developing a test method that is suitable for the many situations where perception of design can provide valuable information seems to be really hard. The product can, for example, be presented in many forms that may all have a different effect on the perception of design. If a product is shown as a picture, the situation is very different from the perception of a 3D object in the participant's hand.

We have focused in our tests and in the development of the test method mostly on the perception of 2D pictures. We have found that motivation is a key element in the perception of design and this is supported also by earlier studies.

The following scenarios could assist a designer in various stages where the visual aspects of product design need to be evaluated. They are by no means aiming at replacing any part of designers' intuitive and creative work. On the contrary, they could enhance the evaluation of alternative designs and the communication within a development group with new views and fresh discussion.

7.1 Scenario 1

One possible future application could involve evaluating whether the intended product claims are perceived by the users. Gaze data could be utilized to define areas and details requiring special attention in a design improvement. The scenario is presented in more detail below.

Research question	Which product details are connected to claimed features and attributes?
Products	Existing product getting re-designed
Test participants	Potential users of the product (10–20 persons)
Analysis methods	Visual heat maps showing attention in different tasks. Statistical gaze data on areas of interest. Statistical data on answers given to evaluation questions and comparison to gaze data.
Possible results	The result can be utilized to define areas requiring special attention in design facelift.

7.2 Scenario 2

Another design improvement approach could investigate which product details (in addition to the logo) carry cues for the brand identity and how well the existing products stand out from the mass of similar products. It is also interesting whether the brand related details are connected to product preference or not.

Research questions	Which product details make the brand recognized? Are these same details used when deciding on preference?
Products	Existing product getting re-designed
Test participants	Potential users of the product (10–20 persons)
Analysis methods	Comparison of gaze data for first impression and preference. Visual heat maps on presumed brand cues and on details affecting preference. Statistical gaze data on areas of interest. Statistical data on preference and comparison to gaze data.
Possible results	The result shows how well existing products stand out from the crowd, which details communicate their brand and how well they do it, and whether these same details play any role in product preference.

7.3 Scenario 3

The approaches above could also be applied in studying new design concept visualizations. Various tasks could be utilized to compare the attention received by the product details to their intended messages. The results would define the initial eye-catching areas and the areas and details which seem to carry the idea of the intended attributes. These areas and details could then be compared to those dominating in product preference.

Research questions	Which concept details jump out? Do the details convey the intended attributes?
Products	Design proposals
Test participants	R&D team (marketing personnel, designers, etc). (10–20 persons)
Analysis methods	Cluster analysis of eye-catching areas. Heat maps and statistical data on areas of interest for different attributes. Statistical data of preference and comparison to gaze data. Effects of different attributes to selection.
Possible results	Results show how well the details of each concept convey the intended attributes. It can also reveal the role of different details and different attributes on final selection. R&D team may also benefit from new views and fresh discussion.

7.4 Scenario 4

A study of competitors’ products or the company’s own older products and their attributes could provide valuable data for the basis of a new product concept. It could reveal means to distinguish the new product from the old ones and underline the importance of some existing product features and their messages.

Research questions	Which details and features stand out from existing products? Do they contain or convey attributes planned for a new product?
Products	Competing products on the market or older products
Test participants	R&D team or potential users (10–20 persons)
Analysis methods	Heat maps and statistical data on areas of interest for different attributes. Statistical data of preference and comparison to gaze data. The effect of attributes and certain details on selection.
Possible results	Results provide data for evaluating planned product attributes and for locating the target design against existing products. They can also reveal the connection between certain details and product preference.

8 Discussion and Conclusions

Using gaze data in the evaluation of perception of design is a methodologically challenging task. There is little previous research in this domain, and much of the project effort has gone into the development of the methodology. As the project ends, this work is still ongoing in continuing basic research projects.

The reason for the complexity of the task is that there are many factors that influence the perception process:

- The presentation form and fidelity of the design product.
- The motivation for the perception task and the way the task is carried out.
- The role of peripheral perception (which could be particularly significant in the case of design products, and which is difficult to measure quantitatively).
- The cognitive processes in evaluation and perception, and their interaction.
- The personal characteristics of the viewer, including education.

At the onset of the project we did not know which of these factors would be most amenable to study using eye-tracking. Therefore the approach chosen was to carry out a series of explorative experiments. Although the research setups became more tightly controlled towards the end of the project, they still involved many factors and several research questions. One clear lesson from our tests was that in order to obtain significant results, the independent variables need to be controlled carefully, and a test set should focus on a single carefully defined question.

Another reason for the difficulty in obtaining a clear breakthrough was that the algorithms for studying gaze data were still rather primitive. We analyzed static variables like fixation durations, number of fixations, and gaze duration in areas of interest. The differences in these static measures were seldom big enough and systematic enough to be statistically significant. From visual observation it seemed that the differences in gaze paths, i.e., the dynamic sequences of fixations, might have been bigger and more informative. However, in the eye tracking literature the algorithms for computing gaze path models from gaze data are only just becoming available as the project ends. We will continue to analyze the data using these new methods.

In addition to developing the research methodology, we also wanted to use it in cases provided by the commercial partners in the project to produce knowledge that would be directly applicable in their design work. We succeeded in this to some extent, though the situation was a bit controversial: the real cases helped in developing the methods, but more could have achieved with more advanced methods, which were now left as a topic for future research.

To summarize, the project dived into a new research area and produced knowledge on the challenges that need to be overcome by further research. Initial interesting results were obtained on the connection of gaze behaviour and subjective preferences, and also on the differences between designers and non-designers (especially in tasks that are cognitively challenging). A wealth of further research topics was identified.

The multidisciplinary co-operation of researchers with eye-tracking background on one hand, and with design background on the other hand, was particularly rewarding.

9 Publications of the Project

Koivunen, K., Kukkonen, S., Lahtinen, S., Rantala, H., and Sharmin, S. (2004). Towards deeper understanding of how people perceive design in products. In: *Proceedings of Computers in Art and Design Education* (CADE2004), Copenhagen Business School, Denmark & Malmö University, Sweden, June-July 2004.

Kukkonen, S. (2005). Exploring eye tracking in design evaluation. In *Proceedings of Joining Forces, International Conference on Design Research*. UIAH, Helsinki, Finland, September 22-24, 2005.

Lahtinen, S. (2005). Which one do you prefer and why? Think aloud! In *Proceedings of Joining Forces*, *International Conference on Design Research*. UIAH, Helsinki, Finland, September 22-24, 2005.

Rantala, H., Koivunen, K., Sharmin, S., and Räihä, K.-J. (2005). Do consumers and cesigners perceive images of design products differently? In *Proceedings of ECVP'05 – 28th European Conference on Visual Perception*, A Coruna, Spain, August 2005.

Räihä, K.-J., Aula, A., Majaranta, P., Rantala, H., and Koivunen, K. (2005). Static visualization of temporal eye-tracking data. In *Human-Computer Interaction* (INTERACT 2005), Rome, Lecture Notes in Computer Science 3585, Springer, September 2005, 946-949.

Sharmin, S., Rantala, H., Koivunen, K. and Räihä, K.-J. (2005) How designers and consumers view a product during memorizing? (Poster abstract) In *Proceedings of ECEM 13 – 13th European Conference on Eye Movements*, Bern, Switzerland, August 2005.

10 References

Abrams, R. A. (1992). Planning and producing saccadic eye movements. In: *Eye Movements and Visual Cognition: Scene Perception and Reading*, Rayner, K. (ed.), Springer: New York, p. 66.

Antes, J. R. (1974). The time course of picture viewing, *Journal of Experimental Psychology*, 103, 62-70.

Babcock, J., Lipps, M., and Pelz, J. B. (2002). How people look at pictures before, during and after image capture: Buswell revisited. In *Proceedings of SPIE, Human Vision and Electronic Imaging*. 4662, 34-47.

Brandt, H. F. (1945). *The Psychology of Seeing*. Philosophical Library, New York.

Buswell, G. T. (1935). *How People Look at Pictures: A Study of The Psychology of Perception in Art*. The University of Chicago Press: Chicago.

Chafe, W. L. (1980). The deployment of consciousness in the production of a narrative. In: *The Pear Stories: Cognitive, Cultural, and Linguistic Aspects of Narrative Production*, W. L. Chafe (ed.), Ablex: Norwood, NJ, 9-50.

Chafe, W. L. (1987). Cognitive constraints on information flow. In: *Coherence and Grounding in Discourse*, Tomlin, R. S. (ed.), Benjamins: Amsterdam, 21-51.

Chafe, W. L. (1996). How consciousness shapes language. *Pragmatics & Cognition*, 4(1), 25-54.

Charness, N., Reingold, E., Pompulin, M., and Stampe, D. (2001). The perceptual aspect of skilled performance in chess: Evidence from eye movements. *Memory & Cognition*, 29(8), 1146-1152.

DeCarlo, D. and Santella, A. (2002). Stylization and abstraction of photographs. In *Proceedings of SIGGRAPH 2002*, 769-776.

Duchowski, A. T. (2003). *Eye Tracking Methodology: Theory and Practice*. Springer: New York.

Ericsson, K. A. (2002). Protocol Analysis and Verbal Reports on Thinking. An updated and extracted version from Ericsson (2002), http://psy.fsu.edu/faculty/ericsson/ericsson.proto.thnk.html

Ericsson, K. A. and Simon, H. A. (1984). *Protocol Analysis. Verbal Reports as Data*. MIT Press: Cambridge, MA.

Ericsson, K. A. and Simon, H. A. (1998). How to study thinking in everyday life: Contrasting think aloud protocols with descriptions and explanations of thinking. *Mind, Culture, & Activity*, 5(3), 178-186.

Hammer, N (1992a). *Möglichkeiten und Grenzen der Überprüfung von Designprodukten durch Okulometrie*. Die Blaue Eule: Essen.

Hammer, N. (1992b). Determining the relevant cognitive models of designproducts by use of eye-movement-recording. In: *Objects and Images*, Vihma, S. (ed.), UIAH: Helsinki, 94-99.

Hammer, N. and Lengyel, S. (1989). Do people perceive what the design expresses? In: *Semantic Visions in Design*, Vihma, S. (ed.), UIAH: Helsinki, pp. m1-m8.

Hammer, N. and Lengyel, S. (1991). Identifying semantics markers in design products: The use of eye-movement recording in industrial design. In: *Oculomotor Control and Cognitive Processes*, Elsevier: New York.

Hayhoe, M. M., Shrivastava, A., Mruczek, R., and Pelz, J. B. (2003). Visual memory and motor planning in a natural task. *Journal of Vision*, 3(1), 49-63.

Henderson, J. M. and Hollingworth, A. (1998). Eye movements during scene viewing: An overview. In: *Eye Guidance in Reading and Scene Perception*, Underwood, G. (ed.), Elsevier: New York, 269-293.

Henderson, J. M., Weeks, P. A., and Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 210-228.

Holsanova, J. (2001). *Picture Viewing and Picture Description: Two Windows on the Mind*. Doctoral dissertation, Lund University Cognitive Studies 83.

Jacob, R. J. K. and Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises (Section commentary). In: *The Mind's Eyes: Cognitive and Applied Aspects of Eye Movements*, J. Hyona, R. Radach, and H. Deubel (eds.), Elsevier: Oxford.

Keinonen, T. (1998). *One-Dimensional Usability: Influence of Usability on Consumers' Product Preference*. UIAH, Helsinki.

Land, M. and Furneaux, S. (1997). The knowledge base of the oculomotor system. *Philosophical Transactions: Biological Sciences*, 352 (1358), The Royal Society, 1231-1239.

Land, M. F. and Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities. *Vision Research*, 41(25-26), 3559-3565.

Law, B., Atkins, M. S., Kirkpatrick, A. E., Lomax, A. J., and Mackenzie, C. L. (2004). Eye gaze patters differentiate novice and experts in virtual laparoscopic surgery training environment. In *Proceedings of the Eye Tracking Research & Applications Symposium* (ETRA04), ACM Press: New York, 41-48.

Liversedge, S. P. and Findlay, J. M. (2000). Saccadic eye movements and cognition. *Trends in Cognitive Sciences*, 4, 6-14.

Loftus, G. R. and Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 565-572.

Lu, S., Olde, B., Cooper, E., and Graesser, A. (2001). Individual differences in reasoning about broken devices: An eye tracking study. In *23rd Annual Conference of the Cognitive Science Society*, Edinburgh, Scotland.

Mackworth, N. H. and Morandi, A. J. (1967). The gaze selects informative details within pictures. *Perception and Psychophysics*, 2, 547-551.

Miall, R. C. and Tchalenko, J. S. (2001). The painters eye movements. *Leonardo*, 34 (1), 35-40.

Molnar, F. (1981). About the role of visual exploration in aesthetics. In: *Advances in Intrinsic Motivation and Aesthetics*, Day, H. I. (ed.), Plenum Press: New York, 385-413.

Morimoto, C. H. and Mimica, M. R. M. (2005). Eye gaze tracking techniques for interactive applications. *Computer Vision and Image Understanding* (CVIU), 98(1), 4-24.

Nodine, C. F., Locher, P. J., and Krupinski, E. A. (1991). The role of formal art training on perception and aesthetics judgment of art composition. *Leonardo*, 26, 219-227.

Noton, D. and Stark, L. (1971a). Eye movements and visual perception. *Scientific American*, 224, 34-43.

Noton, D. and Stark, L. (1971b). Scanpaths in saccadic eye movements while viewing and recognizing patterns. *Vision Research*, 11, 929-942.

Pelz, J. B., Canosa, R. L., Kucharczyk, D., Babcock, J., Silver, A., and Konno, D. (2000). Portable eye tracking: A study of natural eye movements. In: *Human Vision and Electronic Imaging V*, Rogowitz, B. E. and Pappas, T. N. (eds.), SPIE Proc. 3659.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124 (3), 372-422.

Van Zoest, W. and Donk, M. (2004). Bottom-up and top-down control in visual search. *Perception*, 33(8), 927-937.