CHALMERS
UNIVERSITY OF TECHNOLOGY

Eye Tracking as an Evaluation Tool of Perceived Quality in Cars

Master's thesis in Product Development

Mikael Callenberg, Anton Hellåker

Department of Product and Production Development
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2015

REPORT NO. xxxxx/xxxxx

Eye Tracking as an Evaluation Tool of Perceived Quality in Cars

Mikael Callenberg

Anton Hellåker

Department of Product and Production Development

CHALMERS UNIVERSITY OF TECHNOLOGY

GÖTEBORG, SWEDEN 2015

Eye Tracking as an Evaluation Tool of Perceived Quality in Cars

Mikael Callenberg Anton Hellåker

Technical report no XXXXXX

Department of Product and Production Development

Chalmers University of Technology

SE-412 96 Göteborg

Sweden

Telephone   +46 (0)31 772 1000

Gothenburg Sweden 2015

ABSTRACT

The importance of perceived quality within the car industry has been increasing rapidly during the last decade. Identifying and establishing perceived quality requirements is vital to secure Volvo Cars position within the premium car segment.

Volvo Cars has requested new methods to include customers view in the establishment of these requirements. In contrast to objective quality, customers judge perceived quality in a highly subjective way, hard to express and define with words.

This thesis investigates the possibility to use eye tracking as a tool for involving customers view upon perceived quality in cars. The scope is to investigate the maturity of eye tracking in order to design, execute and analyze a study using a sample of 25 employees at Volvo Cars. Two preliminary studies where conducted to understand the abilities and limitations of eye tracking in a customer oriented study. The knowledge gained from process where later implemented in the main study. The main study was designed as a mixed method research, collecting quantitative data using the eye-tracking system Ergoneers Dikablis Essential, followed up by a semi-instructed interview collecting qualitative data. The participants in the study where asked to evaluate the interior quality of a Volvo XC60. The collected eye tracking data where then analyzed in order to find out where people looked the most and first during the evaluation. The qualitative data where then used to understand and interpret *Why* the participants focused on certain areas more than other. Furthermore are correlations between the quantitative and qualitative data identified and interpreted.

The result from eye tracking shows that the *Center Stack* and the *Steering Wheel* are the most viewed areas in the interior of the XC60. 22 out of the 25 participant spent most time evaluating the *center stack,* whereas only 7 participants stated in the interview that this area was the most important and most viewed. The first area evaluated by most participants was according to the observation of eye tracking data the *Steering Wheel*, while only 3 participants stated this in the interview. The complete glance pattern over the evaluated areas in the car is visually presented in heat map, exposing the ratio of time spent on each area of the interior.

The conclusion from this thesis is that eye tracking as a tool can provide valuable information from customers in the area of Perceived quality. It should however be noticed that qualitative data is needed to support the interpretation of the eye tracking data. The main reason for the high attention ratio on the *center stack* and *steering wheel* is according to the authors due to the high complexity in these areas, more focus is therefore needed to evaluate the area. It is also concluded that the participants judge quality from a holistic point of view, which highlights the importance of a harmonized design and execution, where the smallest details can have great impact on the overall impression of the quality.

Keywords: Perceived Quality, Eye Tracking, Mixed Research, Non verbal data collection

PREFACE

This master thesis has been performed at Perceived Quality Geometry & Assurance at Volvo Cars Corporation in Torslanda, Sweden. The thesis was performed in collaboration with the Department of Product Development at Chalmers University of Technology in Gothenburg, Sweden.

The report will reflect the results, conclusions and recommendations resulting from the thesis work.

The thesis work has been including many different areas which resulted in the consultation from many employees in different areas. The authors are most grateful to the assistance and enthusiasm shown from these individuals throughout the work, and without the consultation the thesis work wouldn’t be possible.

A great appreciation is aimed at our examiner Rikard Söderberg which showed on enthusiasm and offered great input though the process.

Other academic researchers which should have thanks for assisting in the thesis work are Konstantinos Stylidis and Dr. Casper Wickman.

We would also like to thank all the members of the Perceived Quality Geometry & Assurance group at Volvo Cars which offered explanatory direction and assistance through the project: Group Leader Martin Hansson together with Emelie Sundqvist, Anders Andersson, Jan Johansson, Lar-Erik Persson, Peter Dahlqvist, Emil Helgesson, Jack Jensen, Tommi Saarento and Elin Schelander.

Other employees at Volvo Cars which has offered assistance and guidance in the project is Reetta Hallila, Volvo Cars HMI-lab, which held introductions of the eye tracking system, and Pernilla Nurbo, Ergonomics Department, which assisted with sample collection.

Finally, we would direct special thanks to our supervisor at Volvo Cars, Dr. Ola Wagersten, for his assistance and time spent throughout the project, and to our academic supervisor, Dr. Karin Forslund, which has provided us with much information and assistance.

All contributors to the thesis work should have many thanks!

CONTENT

TERMS AND ABBREVIATIONS

PQ	The Department of Perceived Quality at Volvo Cars
G&A	The attribute within PQ working with the requirements on Geometry and Appearance
DLO	Daylight opening area, areas which become visible when a door, the trunk or the hood is opened
Ergoneers Group	The company developing the eye tracking system used in the thesis
Dikablis Essential	The name of the specific eye tracking system used in the thesis
CFo	Customer Field of View. The position from where a potential customer will see the vehicle and evaluate its quality.
TP	Test Person. The participants in the Pilot Study and the Main Study are referred to as "Test Persons".
Visual Field	The total field of vision for a human, reaching 220 degrees.
Foveal Area	The smallest, centered area of the visual field. The area with high focus covering 1-2 degrees.
Peripheral Area	The largest area of the visual field. The area with more blurry focus covering 6-220 degrees
Glance	To direct the eyes at or toward something
Blinks	Most often an involuntary act of shutting and opening the eye-lids.
Cross Troughs	A glance passing through an area or object so short it is impossible to determine if the information is interpreted
Corneal Reflection	The reflection of the cornea caused by illumination, moves with respect to the pupil.
Saccade	Quick and rapid eye movements between the fixations
Smooth Pursuit	Slow and smooth eye movements
Fixation on	The paused moment between eye movements fixating the glance an object
Eye Camera	The camera recording the pupil

Field Camera	The camera on the glasses recording the environment
Calibration Point	Point used to calibrate the system, which is used to match the view of the field camera and the eye camera.
D-lab 2.5	Analyze software used to analyze the recordings from the eye tracking system Dikablis Essential.
Marker	Objects in the environment to be visible in the recordings and used to control the drawn AOI's in D-lab 2.5.
Areas of Interest, AOI	Areas drawn within the analyze software D-lab 2.5 to assist in calculations of glance metrics
Glance Metrics	Different values to be calculated in D-lab 2.5, e.g. Total Glance Time and Time to First Glance.
Heat Maps	Drawn maps in gradient colors, from red to green, showing the intensity of glances on the eye tracking recordings

1. INTRODUCTION

This chapter will present an introduction to the master thesis in providing a background to the company, the problem which gives reason to the thesis and to clarify the goal and scope of the project.

1.1 Background

"The importance of the aspects related to the perceived quality will only rise over time. It is happening mainly as a result of the extremely competitive premium automotive sector and drawback of the technocratic way regarding communication of the quality."

- (Stylidis, Wickman, & Söderberg, 2015)

Volvo Cars and the department of Perceived Quality is looking into using eye tracking technology to catch the perception of quality from customers and to see to the potential of establishing an evaluation method for perceived quality in cars.

The increasing competition in the premium automotive sector increases the demands on car manufacturers to differentiate from its competitors, to create a holistic view of quality in its brand and products in order to satisfy its customers' needs and to win market shares from its competitors.

The area of Perceived Quality is an area which has gained more importance over recent years. Assessment of perceived quality is in comparison to objective quality, which can be easily measured, highly subjective and assessed in a higher level of abstraction (Zeithaml, 1988). There is much to gain in mapping and evaluating the factors affecting the perceived quality of costumers and to understand the importance of the holistic view in a design.

In (Mitra & Golder, 2006), the authors states that companies have difficulties in delivering the perception of quality to its customers and that the perception of quality doesn't reflect the actual objective quality of the company's products. In order to be successful and competitive in the premium automotive segment in the future, the automotive manufacturers have to develop and implement perceived quality evaluation methods and techniques (Stylidis et al., 2015).

1.2 Problem Definition

Volvo Cars is currently establishing their requirement of Perceived Quality from sources within the company, through benchmarking of competitors and with the use of experienced employees within the company. A problem when focusing too much on internal knowledge and competitor benchmarking is the loss of input from potential customers.

The problem when including potential customers is that the assessment of Perceived Quality is very subjective and often hard to express in words. The subjectivity within the perception of quality makes it hard to quantify the results as there is no good solution on how to compare the results between participants. This project will look into using eye tracking technology to reduce the need of verbal explanation and to capture customers' perception of quality.

1.3 Volvo Car Group

Volvo Cars is a premium automotive manufacturer founded in 1927 and head quartered in Gothenburg, Sweden. Volvo Cars was bought by Ford Motor Company in 1999 and sold further on to the Chinese automotive manufacturer Geely (Zhejiang Geely Holding Group Co. Ltd.) in 2010. Volvo Cars is a global company with cars sold worldwide and with factories located in Sweden, Belgium, China and Malaysia.

Since mid-2011 Volvo Cars has established a new corporate strategy 'Designed around you'. The strategy is the foundation for the business and the name symbolizes the human-centric brand and company which Volvo Cars wants to be, keeping the customers centered of everything they do (Volvo.Cars, 2014).

The company's vision is to be the world most progressive and desired premium car brand by making life less complicated for people while still improving and strengthen the commitment to safety, quality, and the environment.

The customer is in focus and 'Designed around you' is all about validating and caring for people. See to the customers' needs and to deliver solutions that support them in their daily life and exceed their expectations.

In order to reach the objectives of providing cars that people want, with a top tier premium brand perception, as a lean and nimble company, Volvo Cars need to take actions to close the gap between current state and where they want to be.

1.4 The department of perceived quality

Perceived Quality (PQ) is a sub-division under the Department of Craftsmanship & Ergonomics with responsibility for four different attribute areas with resources allocated according to the importance of the different attributes (see Fig.1).

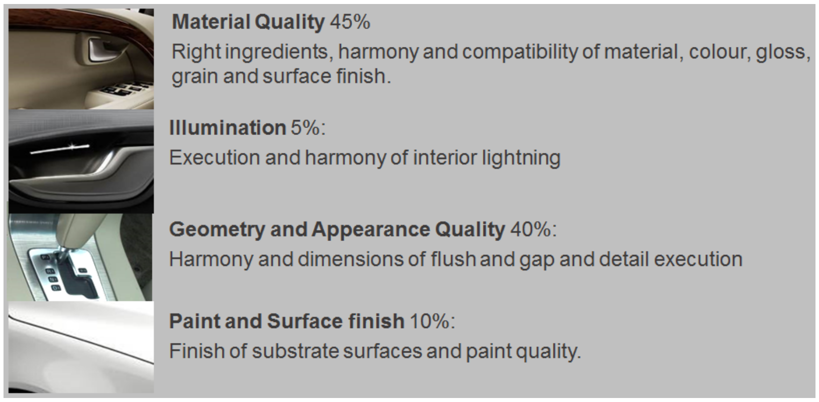


Figure 1: The allocation of resources between the Perceived Quality attributes.

PQ is working to improve the Perceived Quality of Volvo Cars products, to give the products a holistic view of quality to be perceived as premium. It is important to pay attention to every detail that comes into contact with the customers senses to obtain a total harmonious impression.

”Enable an excellent crafted good design, by significantly improve perceived quality of our products to create a sensory delight for our customers, appropriate to the values of our brand”

– The Vision of PQ

Perceived quality is about getting a total harmonious impression for every detail that comes into contact with the customer's senses. To be perceived as premium is to have a good overall quality as well paying attention to every detail, not letting any disturbance or weakness sticking out (PQ.Department, 2012).

The work of PQ is concerning the entire car, with the interior and exterior but also Daylight Opening areas (DLO). DLO areas are the areas which becomes visible when a door, the trunk or the hood is opened.

1.5 Purpose

The purpose of this master thesis is to provide the department of Perceived Quality at Volvo Cars with a tool to assist in the process of establishing and reviewing requirements in perceived quality.

The tool will based on the use of eye tracking technology to investigate where a person looks the *most* when evaluating the perceived quality of a car, and what they are looking at *first* when evaluating the perceived quality of a car.

1.6 Goal

The main goal of the master thesis is to create a structural method for using eye tracking in the evaluation of Perceived Quality and to present results on where people look when evaluating the quality of a car.

1.6.1 Deliverables

The main deliverable is a visual presentation on where people look when evaluating the quality of a car in a way easily used by the employees at the department of Perceived Quality to strengthen their arguments in discussion with other departments.

The second deliverable is recommendations and instructions on how to use the structural method for using eye tracking in the evaluation of perceived quality.

1.7 Scope

In order to reach the goals and to fulfill the purpose of the master thesis there is a need for investigation of the maturity of the technology to see if it's possible to use as a non-verbal method to include customers in the process of establishing requirement on Perceived Quality.

When level of maturity is established there is a need of involving potential customers and to investigate how a customer oriented study using eye tracking could be performed.

Lastly the knowledge gained is adapted in a study to see where customers focus their view the most, and firstly, when evaluating the cars' quality.

The research questions (RQ) presented will serve as a foundation in the project.. Subsequent research questions (SRQ) within each research question adds more depth and a clearer description to the main research questions.

RQ 1: *Is the Dikablis Essential eye tracking technology mature enough to be used as a tool for evaluation of a cars Perceived Quality?*

SRQ 1.1: *What are the problem areas affecting the validity and reliability of a conducted study?*

RQ 2: *How is a customer oriented study of Perceived Quality of cars using eye tracking technology constructed?*

SRQ 2.1: *What are the problem areas affecting the validity and reliability of a conducted study?*

SRQ 2.2: *Which qualitative data collection method should be used to increase the understanding of the eye tracking data?*

RQ 3: *What does a person look at most and first when evaluating the quality of a car?*

SRQ 3.1: *What is the reason to that people look where they look?*

SRQ 3.2: *Are people aware of where they look?*

1.8 Delimitations

The project is ordered by, and carried out in collaboration with the attribute Geometrical & Appearance Quality under the attribute Perceived Quality at Volvo Cars. This means that focus in the project is mainly on geometry and appearance, with the other three attributes secondly prioritized but still considered.

The eye tracking system used in the study is Dikablis Essential developed by Ergoneers Group. This means that the Dikablis Essential system is the only system to be investigated in the project, and no other eye tracking equipment will be taken into consideration in the project.

The Main Study conducted in the thesis is using a sample consisting of Volvo Cars employees. This means that all data, both eye tracking data and the comments and thoughts expressed in interview can't be used for generalization.

The Main Study conducted in the thesis is using one particular car model in the product offer of Volvo Cars as a mediating tool and all the data collected reflects the design and layout of that particular car and cant' be used for generalization.

In the work with Volvo Cars some information is confidential and can't be included into the report. Type of information is clearer description on how the work is performed at the department of Perceived Quality and information regarding the requirements set on the cars.

1.9 Report Outline

Chapter 1: *Introduction* - This chapter explains the background of Volvo Cars engagement in the topic of Perceived quality and the challenge of involving customer in the process of establishing Perceived quality requirements. Furthermore are the scope and the deliverables defined.

Chapter 2: *Methodical Approach* – This chapter describes the methodical process used to answer the stated research questions.

Chapter 3: *Theoretical framework* - This chapter presents published material regarding the main topics of this thesis; Perceived Quality, Eye tracking and mixed methods research. This chapter serves as the backbone of the report.

Chapter 4: *Pre Study* – This chapter investigates and discusses the limitations of the eye tracking system used in the study. Furthermore are suggestions on how to improve these areas presented for forthcoming studies. Last in the chapter is the conclusions made in the Pre Study stated in a bullet point list.

Chapter 5: *Pilot Study* – This chapter presents the design and execution of a small scale study aimed to gain knowledge in how to involve potential customers in the study. Suggestions on how to perform the Main Study are discussed by interpreting observations and collected data. Last in the chapter is the conclusions made in the Pilot Study stated in a bullet point list.

Chapter 6: *Main Study* – This chapter explains the final design, execution and results from the main study. Furthermore are results from the quantitative and qualitative data collection discussed and interpreted to find correlations and explanations to the results.

Chapter 7: *Conclusion* – This chapter presents the conclusions of the thesis based on the stated research questions. The validity and reliability of the study is concluded as well.

Chapter 8: *Recommendations* – This chapter presents the authors suggestions on how further research/work should be conducted in order to establish eye-tracking as a part of the Perceived quality working process.

2. METHODOLOGICAL APPROACH

This chapter gives a short presentation of the methodological approach throughout the master thesis to present a holistic understanding of how the thesis work has been performed. The central steps in the methodical approach are visualized in figure 1.

First step, in order to gain and understanding and to build an informational base within the area of eye tracking, a theoretical framework was summarized. The theoretical framework was compiled through studies of literature within the topic of eye tracking and perceived quality together with an explanation on how the work is performed currently at Volvo Cars. The literature was acquired using databases from Google Scholar and Chalmers library.

The theoretical framework provided information to assist in answering the research questions of the thesis. The methodology was structured by answering one of the three research questions separately in three different parts of the thesis work, *Pre Study*, *Pilot Study*, and the *Main Study.*

The Pre Study is performed in order to answer the first research question if Dikablis Essential eye tracking technology is mature enough to be used for evaluation of Perceived Quality in the development of cars. The methodology is to get more acquainted with the system through investigational use in connection with a car.

The Pilot Study is performed in order to answer the second research question of how the potential customers can be involved into the evaluation using eye tracking technology. The methodology is to involve associates at Volvo Cars to simulate potential customers

The Main Study is performed in order to conduct a study with the knowledge from the two part studies and to analyze the results to establish possible conclusions and recommendations made to Volvo Cars.

The methodology used, the results obtained, and the discussions held during the work of each study will be reported in the report.

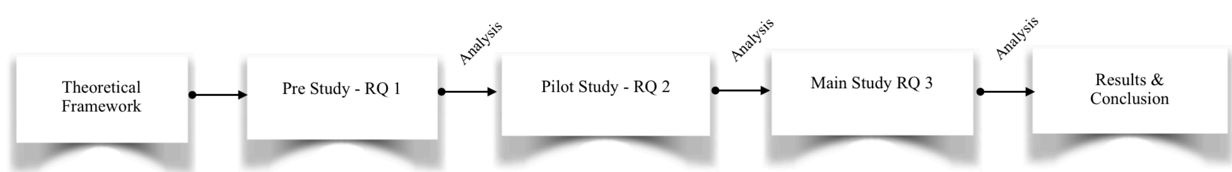


Figure 1. Central steps in the thesis process showed in chronological order.

3. THEORETICHAL FRAMEWORK

In this chapter follows an overlook of the sub chapters within the theoretical framework serving as a knowledge base for the thesis.

3.1 - Theory of perceived quality. Investigation was made to understand the definition and view on perceived quality. What are the important factors when analyzing perceived quality?

3.2 - The current of work at PQ. In this section is a description of the procedure of work at the department of Perceived Quality at Volvo Cars, focusing on the area of Geometry and Appearance. How are the requirements set today and where are information gathered?

3.3 - Eye tracking technology. The third topic is information regarding eye tracking. The information includes the anatomic behind the technology, how the actual technology works and in which applications it has been used earlier.

3.4 - The eye tracking system used. The fourth topic is to give more understanding in how the eye tracking system used in the project works, including how the system works and the terminology used when talking about the system.

3.5 - Eye theory vocabulary list. The fifth section is a summarizing vocabulary list including terminology used within the area of eye tracking to assist in the reading of the report.

3.6 - Designing a research study. The sixth topic offers knowledge in how to conduct a research study in order to achieve measurable, comparable and statistically significant data. Areas including qualitative, quantitative and mixed methods, as well as validity, reliability and sample has being reviewed in order to identify an appropriate strategy for the study.

The information resulting in the theoretical framework was gathered through literature reviews together with internal and external contact sources.

3.1. Perceived Quality

The concept of quality can be analyzed from two different perspectives according to (Espejel & Fandos, 2008), Objective quality referring to aspects of the product that is measurable and verifiable technically, while the perceived (subjective) quality refers to the consumers' perception and individual judgment of the product. (Mitra & Golder, 2006) add that perceived quality is judged relative to the quality that is expected of the product and (Garvin, 1984) states that the perception of quality can be as subjective as the assessment of aesthetics.

(Zeithaml, 1988), states that Perceived Quality differs from that of objective or actual quality and includes a global assessment of a product in a higher level of abstraction, and that the judgment of Perceived Quality usually is made within the consumer's evoked set. The close relation to aesthetic evaluation is highlighted through research and what they have most in common is stated by (Maxfield, Dew, Zhao, Juster, & Fitchie, 2002), which argues that this kind of quality is perceived through visual inspection and comparison. These theories lifts that Perceived Quality is affected by earlier experiences and that evaluations of quality is made in a comparison context (Zeithaml, 1988).

According to (Mitra & Golder, 2006), companies have difficulties in giving the perception of quality to the consumer and that customer perceptions of quality do not reflect the actual objective quality the companies offer. And this while (Mitra & Golder, 2006) claim that it is the perception of quality, and not the objective quality, that result in good sales, profitability and satisfaction. These theories and assumptions gives some prove that Perceived Quality is of great importance for companies.

3.1.1 The Definition of Perceived Quality

The definition of Perceived Quality is an area treated by many and the definition used differs between researchers. To be able to handle the wide area of Perceived Quality it is suggested to further dividing it.

(Espejel & Fandos, 2008) discusses the need of further dividing perceived quality into two factors: intrinsic and extrinsic attributes. This factorization is reinforced with several different articles and is seen as a relevant classification of the concept Perceived Quality. Intrinsic attributes are related to the physical aspects of the product, e.g. color, shape, appearance etc. while extrinsic attributes are related to the non-physical aspects, e.g. brand, place of origin, price etc. (Espejel & Fandos, 2008).

A similar definition of the concept is suggested by (Stylidis et al., 2015). The authors suggest the definition of Value Based Perceived Quality (VPQ) including both the extrinsic and intrinsic quality. The extrinsic quality in VPQ is called external factors (e.g. branding and customer behavior) and the intrinsic quality is called Technical Perceived Quality (TPQ) and it represents the engineer approach (see Fig.2).

Technical Perceived Quality can be divided into three major components; Visual- Feel- and Sound Quality. These areas can be further divided into sub components and the wide spectrum of components including in the concept of Perceived Quality demands delimitations in evaluation.

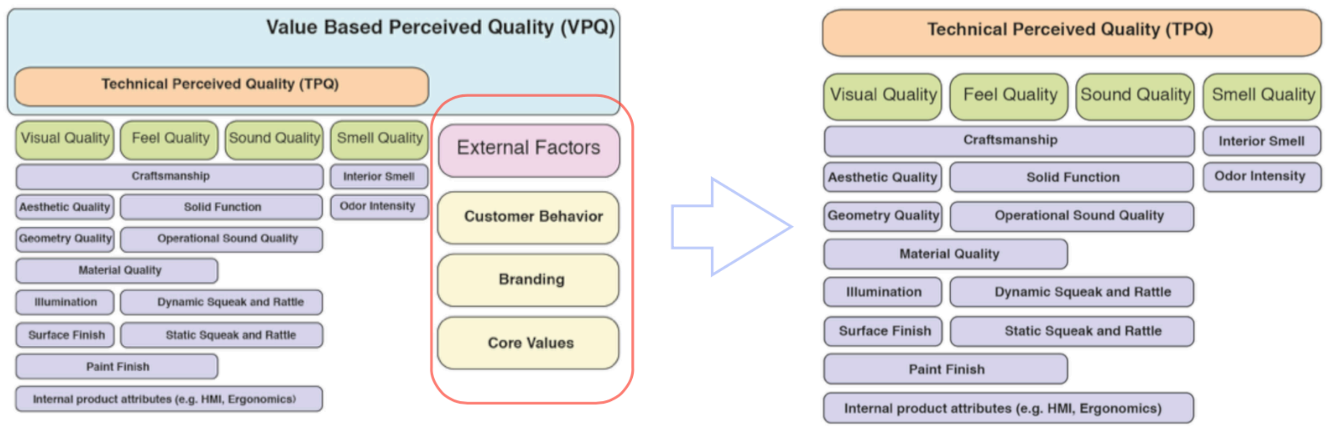


Figure 2: The Value Based Perceived Quality with focus on the intrinsic quality called Technical Perceived Quality explained by (Stylidis et al., 2015).

3.1.2 Perceived quality is gained over time

Perceived Quality is something that a company builds over time and that often results in loyalty from the customers and an increasing brand reputation. One example is the search engine Google and its success on its market. It started as a small company competing with many other search engines but is now the number one search engine worldwide with 88,5% of all searches during the period Jan 2014 to Jul 2014 (see Fig.3). Some people may also use the word “Google” as a synonym to “Searching the web”

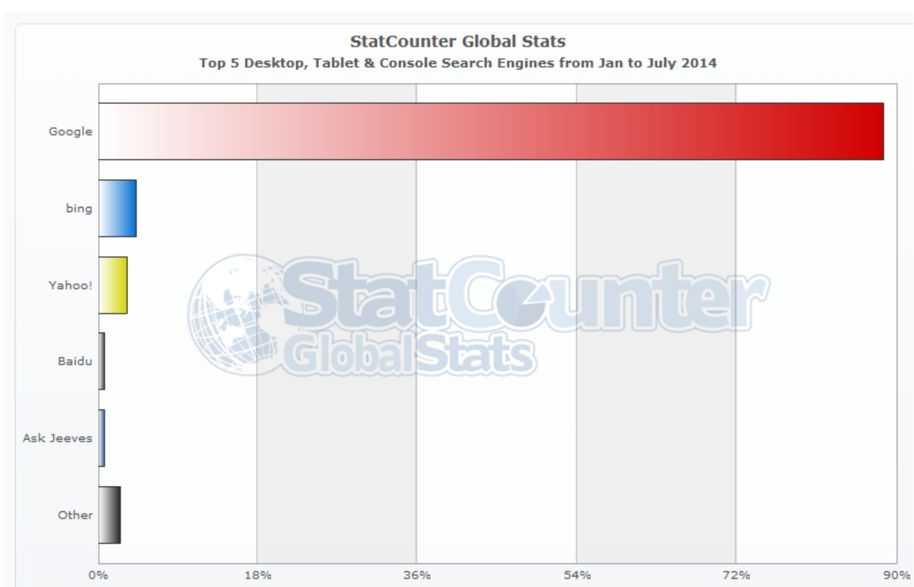


Figure 3: Graph showing the percentage of internet searches performed with each search engine.

(Mitra & Golder, 2006) expects asymmetric changes in quality between increases and decreases, and that customer's perception of quality decreases more with decreasing objective quality than it would increase with an equivalent increase of objective quality. It is easier as a customer to lose confidence in a company's quality than to gain confidence. This is true both in short-term and long term effect according to (Mitra & Golder, 2006).

3.1.3 Perceived Quality is judged in a comparison context

It is suggested that Perceived Quality is judged in a comparison context with other similar products, and the quality is according to (Maynes, 1976), evaluated as high or low relative to the products viewed by the consumers. (Zeithaml, 1988) highlights the importance of the consumer and states that it is the assessment of competing products done by the consumer's, not the company, which matters.

In today's market, where there are many different brands and products providing the same solutions, it is needed to not only differentiate your product in the market but also in the desired segment of the market were your product are supposed to act.

Perceived Quality can differentiate a brand from other brands with the equal objective quality. A study conducted by (Clodfelter & Fowler, 2001), showed that consumers evaluating the quality of apparel is affected by the extrinsic attributes (e.g. place of origin and brand).

Another example is the automotive market, which is filled with cars of every size and performance. This meaning that if you want to be perceived as a premium brand, then the perceived quality of you products needs to be compared to other brands perceived as premium. (Stylidis et al., 2015) states that "Perceived quality is one of the most important factors underlying success of car manufacturers today".

Perception of Quality is also strongly connected to price. If a product is perceived as of great quality, customers are ready to spend more on that product than others of equal objective quality. According to (Clodfelter & Fowler, 2001), it is showed that consumers use price to judge on quality when it is the only cue.

The familiarity of a brand is highly affecting the perceived quality of its products and comparing a brand in context with earlier owned cars affect the perception of quality. In a study performed by ALG (alg, 2012) it was clear that a person who himself owned a car of a certain brand perceived that brand to be of higher quality than a person who did not own a car of that brand (see Fig.4). For example was the Mazda-owners perception of the quality of Mazda as high as 81, while the non-owners perception of Mazda was 64.

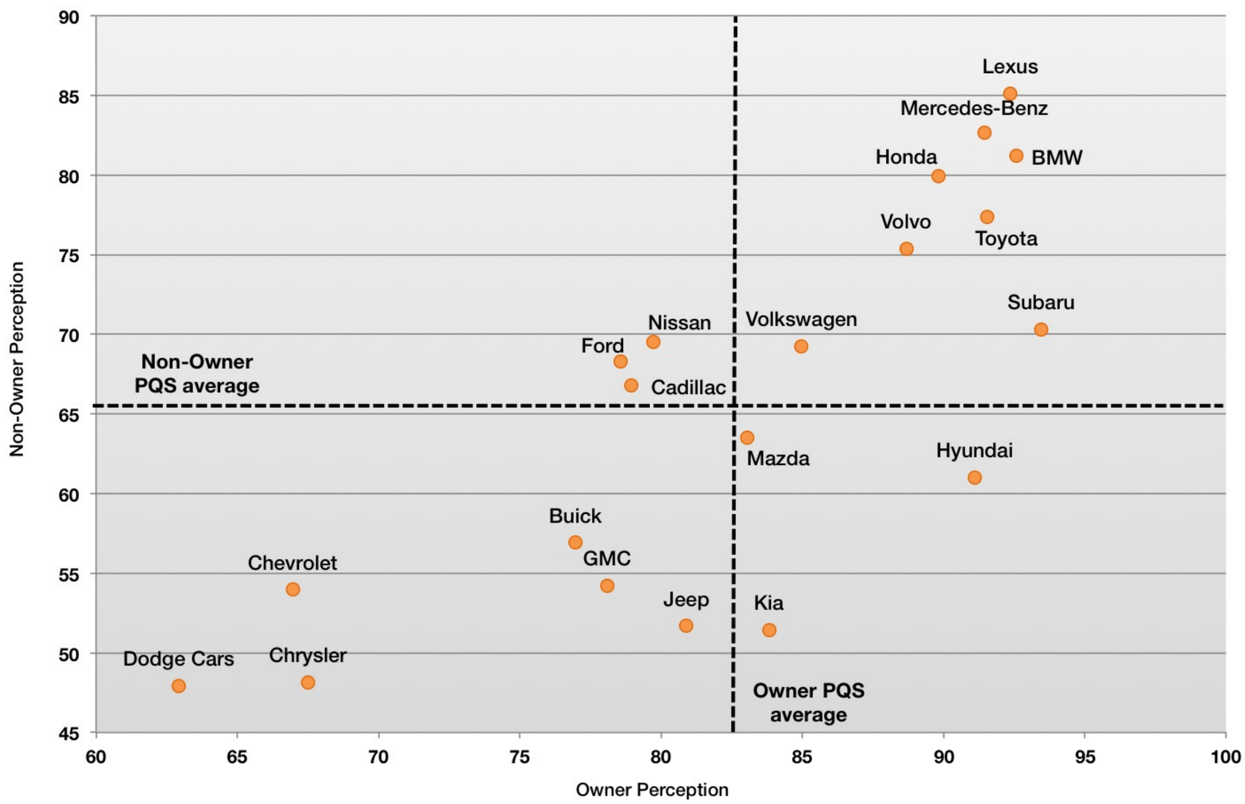


Figure 4: Chart of perception of quality of cars between owners and non-owners (alg, 2012).

This theory that earlier used products affects perception is also stated by (Baxter), which divides attractiveness towards consumers into four different aspects affecting the attraction, (1) recognition to earlier used products (2) apparent functional attraction, differentiated from the real function (3) symbolic attraction affected by the social and personal values of the consumer (4) and an intrinsic attraction to the inherent beauty of the product.

3.2 The current work procedure at PQ

The current procedure of work at PQ at Volvo Cars consists of establishing requirements on perceived quality related areas of the entire product. The requirements are set on the base of earlier experiences and own knowledge but the main source of information is through benchmarking of competitors.

Benchmarking is important as the number of different players within the premium segment in the automotive industry is becoming limited (Stylidis et al., 2015). This leads to that competitiveness are to be seen as one of the important dimensions of perceived quality and that assessment and evaluation of perceived quality attributes is highly affected by the competitors.

"It is possible to say that the level of the perceived quality in the premium segment is driven to a great extent by the competition among the players."

- (Stylidis et al., 2015)

3.2.1 The method of performing a benchmark at Perceived Quality Department

The competitors for each project at Volvo Cars are decided by the project administration and the competing cars relevant to benchmark are ordered to PQ. The benchmark is performed with the overall objective to evaluate how Volvo Cars is standing against its competitors and to see competitor's solutions on perceived quality related problems.

The Benchmarks helps to improve knowledge about Perceived Quality and to lift new possible requirements, as well as it supports the process of the project with clear documentation on how the competitors are performing. The benchmark may show on new solutions but also the requirements needed to match the competition.

Each attribute within PQ are responsible for the benchmark on its respective areas, e.g. G&A measures and analyzes the Geometry and Appearance of the competitors.

3.2.2 Requirements set by Geometrical Appearance Quality (G&A)

The requirements on Geometry and Appearance are set on the entire car. Geometrical requirements concerns the appearance of gaps, radii's, and ball corners, and these requirements are measured in mm. Appearance Requirements are set on the entire vehicle involving the appearance of each component and their interrelation with each other.

The appearance requirements are set to secure premium execution of the details that are visual for customers. A list of parameters is evaluated to ensure that the final solution has premium Perceived Quality. The parameters are evaluated based on whether they are visible from the Customer Field of View (CFoV), the angles and positions defined on seating positions, relevance and customer heights.

After establishing the requirements early in the process, these requirements need to be reviewed and controlled throughout the development of the car. The reviews of the car in development are done both through digital visualizations but also physically in investigating the different stages of pre-built cars. The different reviews during the process have different objectives with different level of detail in the investigation but are all performed to assure that the requirements of PQ are met.

3.3 Description of eye tracking technology

Eye tracking methods has been developed for more than 100 years, where the early systems relied on contact lenses that covered both the cornea and sclera. A metal coil was embedded in the outer edge of the lenses to capture fluctuations in the electric field when the eye was moving (Duchowski, 2007). These methods where seen as quite invasive for the person participating, which is why present eye tracking solutions relies on techniques without physical elements in contact with the eyes. The technology is continuously developing and this theory chapter aims at giving a description, with focus on the important concepts in the field of eye tracking.

3.3.1 The Anatomy of the Eye

The human eye is a very complex organ consisting of many parts working together (see Fig.5).

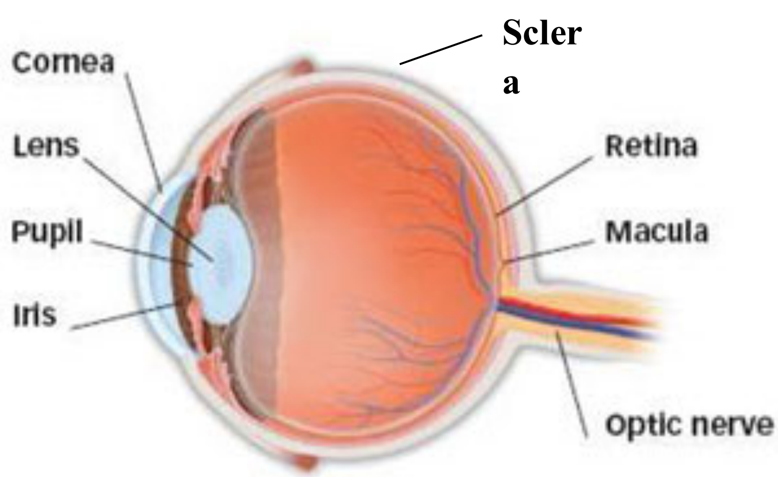


Figure 5: The anatomic structure of the eye including the main parts in the process of seeing (GetEyeSmart, 2015).

The function of the eye is to collect information regarding the surrounding environment and transform that information to the brain for interpretation (Segre, 2014). Light from the environment is focused through the cornea and the lens onto the retina at the back of the eye, which transforms the optical images into electronic signals that can be transported to the brain.

The different anatomic parts are described in table 1.

Table 1: **Descriptions of the main anatomic functions in the process of seeing**

Anatomic Function	Description
Cornea	The clear outer surface of the eye letting light into the eye.
Lens	A spherical body focusing the light rays onto the retina.
Pupil	The center of the eye which opens and closes to regulate the amount of light passing through to the retina.
Iris	A pigmented membrane controlling the opening of the pupil.
Sclera	Outer layer of the eyeball surrounding the eye forming the visible white of the eye.
Retina	A light-sensitive inner lining transforming the optical images from the lens into electrical signals.
Macula	Placed near the middle of the retina and allows for objects to be viewed with great detail.
Optic Nerve	The nerve which carries the electrical signal to the brain for interpretation.

3.3.2 Visual field

There is an enormous amount of visual information available that affects us at all times. The human eye is monitoring a visual field of up to 220°, but can only receive detailed information of 2°, this smaller area of high focus is called the Foveal area.

The Para-foveal area of 2-5° just outside the Foveal area (see Fig.6) is an area of lesser focus than the Foveal area but still better than the peripheral area which reaches from 6-220°. The objects in the visual field are most clear and colorful in the Foveal area and gradually gets more blurry and colorless towards the edges of the visual field (Tobii, 2015).

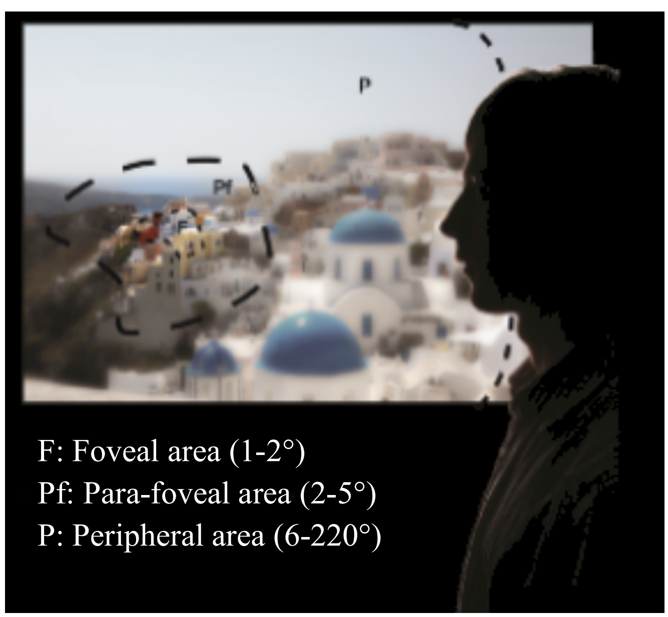


Figure 6: The different area of focus in the visual field of the human eye (Tobii, 2015).

3.3.3 Blinks

Blinking is an essential function of the eye, rapidly providing the eyeball with lube to prevent dehydration. The blink speed is highly individual and can be affected by elements such as fatigue, decease and medication. Studies have showed that the mean blink duration of a healthy human is between 100- 150 Milliseconds. (UCL, 2006)

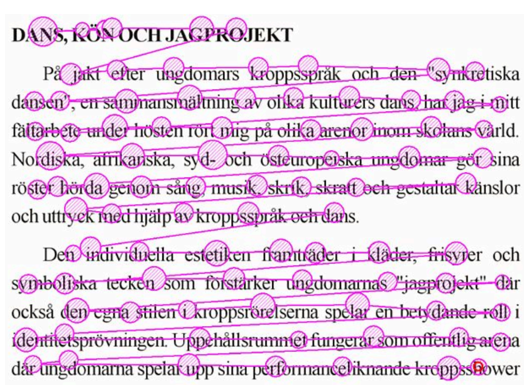
The blinking rate is controlled by the blinking center but can also be initiated externally. The blink rate is hugely affected by the tasks the evaluated person is engaged in. (Bentivoglio et al., 1997) showed that the mean blink rate of 150 healthy males and females during a conversation was 26 times / minute, whereas the rate during reading was as low as 4,5times/minute during reading.

3.3.4 Eye Movements

Eye movements are made to always keep the image on the retina to get focus even in the case of movement of one's head (Tobii, 2015). According to (Kowler, 2011) is the image motion on the retina crucial for vision, where both too much and too little motion can be a problem. Too much motion will mean that the eye don't have the time to focus on the object while too little motion may result in objects fading from view within seconds.

The small area of high focus requires quick and large scanning movement of the eye to get new things in focus. These scanning movements are called saccades, during the history a vast amount of research has been made to identify how often these saccades occur. (Richardson & Spivey, 2004) claims 3-4 times per second, whereas (Collewijn & Kowler, 2008) research shows a maximum 2-3 times per second.

When the eye is focusing on an object and pauses from the saccadic movement, it is called fixation, and it is the fixations and saccades that are responsible for getting objects of interest into the Foveal area for maximum focus. Figure 7 shows the collaboration between saccades and fixations when a person read a text.


DANS, KÖN OCH JAGPROJEKT

På jakt efter ungdomars kroppsspråk och den "synkretiska dansen", en sammansmältning av olika kulturers dans, har jag i mitt fältarbete under hösten rört mig på olika arenor inom skolans värld. Nordiska, afrikanska, syd- och östeuropeiska ungdomar gör sina röster hörda genom sång, musik, skrik, skratt och gestaltar känslor och uttryck med hjälp av kroppsspråk och dans.

Den individuella estetiken framträder i kläder, frisyrer och symboliska tecken som förstärker ungdomarnas "jagprojekt" där också den egna stilen i kroppsrörelserna spelar en betydande roll i identitetsprövningen. Uppehållsrummet fungerar som offentlig arena där ungdomarna spelar upp sina performanceliknande kroppsshower

Figure 7: Fixations marked with a circle and lines represent the Saccadic movement (Wikipedia, 2015)

Another Eye movement is the Smooth Pursuit which is a smoother tracking of targets comparing to Saccades. Smooth pursuit can’t be voluntarily controlled as it isn’t possible to initiate a smooth pursuit in a stationary environment or to completely retain it in an environment with moving objects (Kowler, 2011).

Scanning environments with multiple targets involves a complex collaboration between saccades and smooth pursuits, however research has found that the same neural mechanism is present prior the selection of target regardless if the target is detected by a saccade or a smooth pursuit (Krauzlis & Dill, 2002).

3.3.5 How Eye tracking technology work

Today, most of the commercial eye tracking system takes advantage of the corneal reflection, which is a reflection resulting from using an infra-red light source pointed towards the targets eye. Infra-red light is used in order to prevent the eye from dazzling which occurs using a visible light source.

When infra-red light is aimed towards the retina; a large proportion of this light is reflected back, triggering the pupil to appear as a well-defined bright disc (Milekic, 2004). This effect only appears if the light source is coaxial with the optical path and is often referred to as "bright pupil technology". Figure 8 shows the similarity between the bright pupil effects compared with the "red-eye" effect caused by a compact camera using a flash close to the camera lens in low ambient light.

The illumination of the eye serves another important function, it generates the corneal reflection. This reflection, also known as the "First Purkinje image" or the "Glint" appears as a small but sharp area outside the pupil (See Fig.8) (Goldberg & Wichansky, 2002). The corneal reflection will appear at different places of the eyes with regards to the angle of the gaze vector.

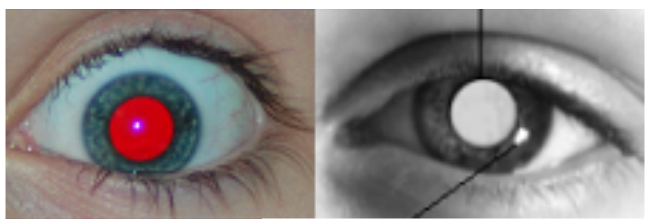


Figure 8: To the left, red-eye caused by compact camera. To the right, the effect of "bright – eye" (Goldberg & Wichansky, 2002).

The eye tracking system needs to detect both the center of the pupil and the corneal reflection and subsequently measure the vector between them. Since the anatomy of humans differs it is necessary to perform a calibration process for every individual using the system. This calibration works by displaying one point in the environment captured both by the eye and the field camera. When the gaze is aimed towards the point it is manually marked in the recording from the field camera. The system records both the pupil and corneal-reflections relationship to a specific x and y coordinate on the field cameras recording. This procedure is often repeated in order to increase the precision (Goldberg & Wichansky, 2002).

3.3.6 The applications of eye tracking

Several fields have showed interest in the eye tracking technology during the history. In 1970 scientists discovered that people suffering from praecox dementia (Schizophrenia) showed difficulties in performing smooth pursuits when following a moving object, instead an elevated use of saccades where used (Holzman et al., 1974).

Another field where eye tracking has been studied in great depth is the process of reading. Studies performed by (Pollatsek, Rayner, & Collins, 1984) revealed that the average person fixate on a 4-8 letters word for approximately 200-250 milliseconds and words of 2-3 letters are being skipped 75% of the time. The first studies conducted towards consumer behavior where conducted in the forties in order to understand what factors that determine the amount of attention assigned to different advertisements (Richardson & Spivey, 2004).

Later research conducted with eye tracking has focused on ergonomic design and computer interfaces, the biggest benefits of eye tracking are the possibility to improve user interfaces and to understand the human behavior (Tobii, 2015). The human brain automatically steers the eye to objects in its field of view that is of interest, and this shows which information that's processed by the brain. Test-persons have in general a hard time explaining what they observing and what particular details that falls under interest. Eye tracking assists in noticing what a person involuntarily thinks about as it is suggested that people look at what they think about (Reid, MacDonald, & Du, 2013).

The eyes are a very efficient tool for pointing and are used in the daily life within interaction with the world around us. Eye tracking enables to use a person's gaze as a fast, intuitive, and natural way of interacting with computers (Tobii, 2015).

Typical measurements that can be obtained with an eye tracker are: (1) Fixations, to see when the eye is stabilized over an area/object of interest, (2) Fixation Time, measurement of the duration of fixation on that area/object, and (3) Time to first glance, measurement on how fast a specific area of interest is fixated on (Reid et al., 2013).

Several attempts have been made to decode how the different measurements can be interpreted. (Poole, Ball, & Phillips, 2005) argue that more fixations on a specific area of interest indicate that the area is more important or noticeable than other areas in the visual field. It can also depend on the targets complexity and is more difficult to encode (Jacob & Karn, 2003). A longer fixation does according to (Just & Carpenter, 1976) indicate difficulty to extract information, but could also mean that the object is more engaging in some way. (Byrne, Anderson, Douglass, & Matessa, 1999) interpret the measurement "time to first glance" as the area has better attention properties compared to other areas in the visual field.

Although various of explanations of how eye-tracking metrics should be decoded there are according to (Poole & Ball, 2006) still a need of further research to be conducted in order to standardize how the metrics should be used and interpreted.

3.4 The Eye tracking System used in the project

This chapter will provide a shorter description of how to work with the system and for a deeper description on how the system is operated and how the analyze tool D-lab 2.5 works it is suggested to read the Dikablis Essential Handbook (Ergoneers.Group, 2011) and the D-lab 2.5 Manual (Ergoneers.Group, 2012).

3.4.1 Ergoneers Group

The Eye tracking system used in the project is manufactured and developed by the Dutch company Ergoneers Group and is named Dikablis Essential. The Dikablis Essential system is the first Eye tracking system developed by Ergoneers Group and was released on the market at 2009. Later two other systems were developed with one, Dikablis Professional, released during the fall of 2014 and one, Dikablis Enterprise, to be released during the first quartile of 2015.

Ergoneers Group was founded in 2005 as a spin-off from the department of Ergonomics at the Technical University of Munich. It is today located in Manching and Geretsried in Germany as well as Portland, USA. Ergoneers Group develops, manufactures and sells simulation, measuring and analysis systems for research and optimization of the interaction between people and machines (Ergoneers, 2015).

3.4.2 Adjusting and calibrating the Dikablis Essential

Dikablis Essential consists of a pair of glasses attached on the head equipped with two cameras. One camera, referred to as the “Eye Camera” is filming the eye and one camera, referred to as the “Field Camera”, is filming the environment. These two cameras are together filming and locating where a person’s pupil is located to see where he/she is looking.

First step in using Dikablis Essential is to adjust and calibrate the system to the head and eyes(Ergoneers, 2011). The steps of adjustment and calibration are needed to be performed for every new person, and preferably controlled between recordings to assure a statistically reliable collection of data.

Adjustment of the cameras is manually performed to fit each person’s eye and head. First step is to adjust the eye camera manually in a way that the eye is centered in the image as horizontal as possible (see Fig.9).

Figure 9: Optimal horizontal position of the eye (Ergoneers.Group, 2011).

The distance between the eye camera and the eye needs to be adjusted to get a sharper image of the eye. The eye camera may be needed to adjust vertically if a person has long eyelashes or a lot of make-up (see Fig.10). The focus of the field camera needs to be calibrated to gain a good recording of the environment and this is done manually by rotating the lens of the "field camera" (see Fig. 10).

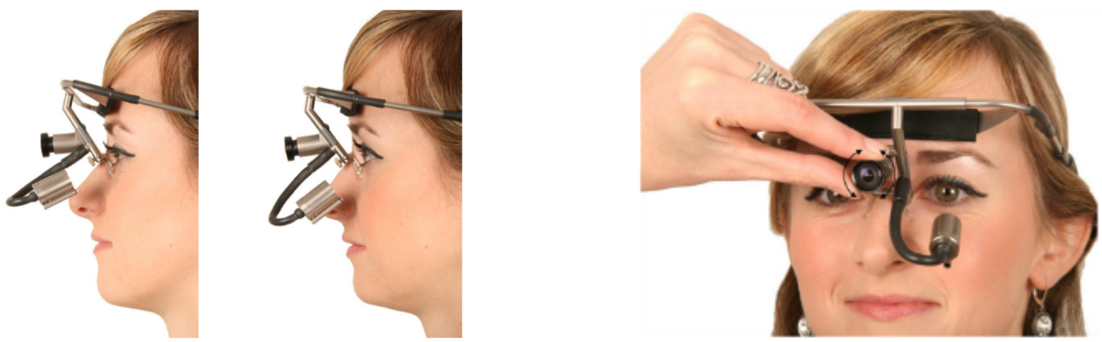


Figure 10: To the left, vertical adjustment of the eye camera. To the right, adjusting focus of the lens (Ergoneers.Group, 2011).

When the adjustments of the cameras have been made, the next step is to use the 'Calibration Wizard' to calibrate the system. The calibration wizard will guide through three steps to give support in finding the optimum detection of the pupil but will still demand a lot of interaction and controlling by the person performing the calibration.

The detection of the pupil is effected by how well it can be isolated from other dark areas around the eye (e.g. eye-lashes, make-up and shadows). This is manually set and when the detection of the pupil is satisfying, the calibration step can be performed.

The calibration is done through that the person equipped with the glasses looks at four different points while keeping its head still and these four points are to be marked on the screen with the cursor. During the calibration the video of the eye is visible to see if the detection of the pupil (the green circle) remains good throughout the calibration (see Fig.11).

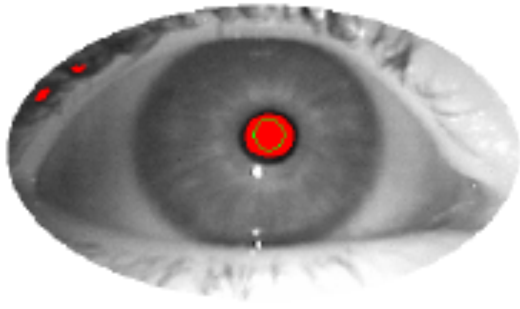


Figure 11: The green circle indicating that the pupil is detected (Ergoneers.Group, 2011).

The four different calibration points are one by one marked on the screen while the person equipped with the glasses focuses on them without moving its head. A green marker will appear at the point and the new quadrant to calibrate in will appear (see Fig.12).

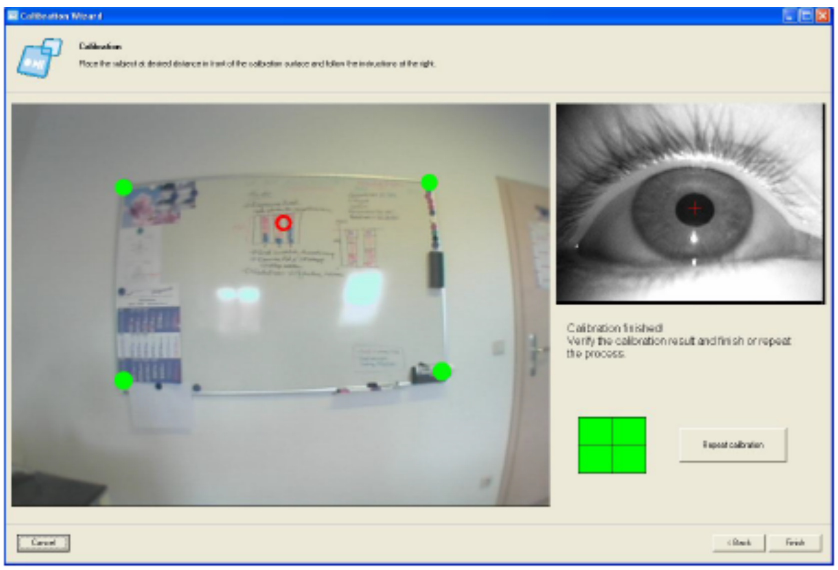


Figure 12: All four calibration points displayed with green markers (Ergoneers.Group, 2011).

3.4.3 Using Markers

In order to analyze the data recorded with the eye tracking system markers are needed in the environment. The markers mark reference points in the surroundings and help to draw AOI's in a stable way and to assure that heat maps remain even though the head is moving (Ergoneers.Group, 2012).

With Dikablis Essential, 16 different markers are delivered to be used to assist in the analysis of Eye tracking videos. The markers is a combination of white symbols on a black squared background (see Fig.13), the markers are named after big worldly cities. The total collection of the 16 markers available can be viewed in appendix 1.

Figure 13: An example of the markers used in Dikablis Essential (Ergoneers.Group, 2011).

To assure a good evaluation of the data the markers needs to be positioned correctly and some guidelines are provided by Ergoneers Group (Ergoneers.Group, 2012).

- Always position the markers as near as possible to the area of interest, AOI. If possible, the markers and the AOI should be on the same plane.
- Make sure that the entire marker is in the picture. Markers which are not entirely visible will not be detected.
- The lighting should be set so that the contrast in the marker pattern is as high as possible. Ideal lighting conditions are bright and shady locations or artificial light.
- If possible, position the markers so that the camera is aimed at them vertically. This will greatly increase the probability that they will be detected. Even so, markers can also be detected if the camera is pointing towards them at an angle. In such case, make sure that the marker resolution is good, i.e. the markers are large enough.
- Make sure that the resolution of the markers is large enough. The size of the marker which is necessary for optimal detection is greatly influenced by the lighting conditions, the distance between the glasses and the markers and the viewing angle.
- The markers orientation is irrelevant. They will be detected regardless of their rotation.

3.4.4 Analyzing eye tracking data using D-lab

D-lab 2.5 offers several functions to analyze and visualize the eye tracking data. The following functionalities will be explained in general:

- Drawings of Area of Interests (AOI)
- Elimination of Blinks and Cross-troughs
- Glance Metrics
- Heat maps

Draw Area of interest (AOI)

The areas of interest is the foundation in analyze of the eye tracking data, within these defined areas the software is capable of calculate the glance metrics. The object or area of interest is defined in D-lab by identifying a frame from the field camera where the object and the corresponding marker is visible (see Fig.14). The lines containing the area can be designed to include any shapes in two dimensions.

Figure 14: AOI's and markers in the interior of the car (Printscreen from D-lab 2.5.

When the desired AOI's are drawn within the software it is time to calculate glances. This is done by pressing the button calculate glances. Now all glances within each and every of the AOI's is displayed in the time frames in the time line beneath the eye tracking video (see Fig.15).

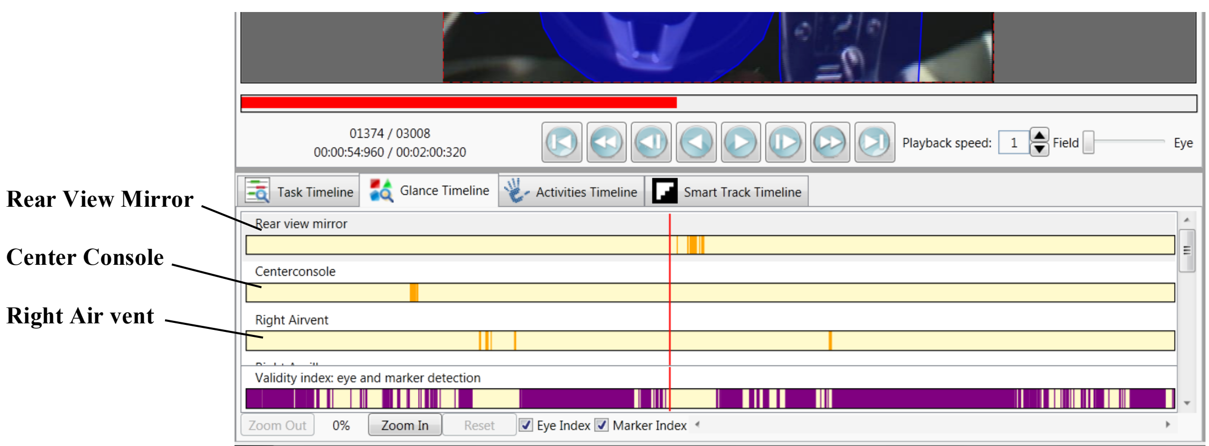


Figure 15: The timeframes under the eye tracking video showing all the AOI's and the connected glances (Printscreen from D-lab 2.5).

Elimination of blinks and cross troughs

When the glances of each AOI have been calculated the recording needs to be prepared for further calculations.

Due to the physics of the human body there is a need of adapting the raw eye tracking data obtained in the recordings (Ergoneers.Group, 2012). During a recording the test-person wearing the eye tracking system will blink. These blinks will be interpreted by the system as a short disruption when the system lose track of what the test-person glances towards. When analyzing the data these interruptions will be interpreted as if the test-person glance towards an Area of interest (AOI), and be stopped when the blink is initiated.

This will lead to that longer glances towards the same AOI is fragmented and counted as several short glances instead of one longer. To avoid this phenomenon the supplier of the system recommends that interruptions in the recordings shorter than 120ms, which correspond to the length of a blink, is being removed. The software then automatically interweaves the interrupted recording and joins together the previous and subsequent glances.

A Cross-trough is a very short glance towards an AOI that the system cannot interpret as if the test-person actually looked at the area or just passed through. For instance, when a test-person sweeps from the Ceiling of the car to the Center stack, the instrument panel is passed through. However, since the time glanced towards the instrument panel is very short it is not sure if the test-person actually intends to look there. To eliminate these glances to be counted D-lab offers the possibility to establish a time limit. The supplier suggests setting this value to 120ms, meaning that glances towards an AOI under 120ms will not be counted during the calculation of glance metrics.

Glance metrics

When elimination of blinks and cross troughs has been made, glance metrics from the eye tracking data can be calculated.

Glance Metrics is calculated in D-lab and there are several of “build in” metrics available to be calculated, some examples on glance metrics available is seen in table 2, and all metrics can be seen in the manual (Ergoneers.Group, 2012). When performing the calculation of glance metrics, D-lab is going thru the recordings frame-by-frame to identify where each glance is located at the particular time. The calculated glance metrics are displayed in a table within D-lab 2.5 that can be exported to other software’s for further work.

Table 2: **Examples of glance metrics possible to calculate with D-lab 2.5**

Glance Metric	Abbreviation	Unit	Explanation stated in D-lab
Total glance time	GTOT	Sec	Accumulated duration of glances towards the AOI for the selected time interval in seconds.
Number of glances	GNUM	No	Number of glances to the AOI for the selected time interval.
Mean Glance Duration	GMEAN		Average glance duration to the AOI for the selected time interval in seconds.
Attention ratio	ARATIO	Percent	Ratio representing the percent of time when glances are within the selected area of interest during the selected recording.
Time to First Glance	TFG	Sec	Duration in seconds from the beginning of the recording to the first glance upon the specified AOI.

Heat Maps

A heat map is according to (BusinessDictionary, 2015) defined as a "graphical representation of data using colors to indicate the level of activity." A more intense color often indicates higher activity, and the scale used in D-lab 2.5 is moving from green (low activity), through yellow (medium activity) until red (high activity) (see Fig.16). Heat maps are well suited when to visually display large amount of data and to identify similar values (Custom-Analytics, 2014).

Figure 16: Picture displaying a Heat Map indicating glance activity in gradient color from green to red (Ergoneers.Group, 2013).

3.5 Designing a Research Study

This chapter will provide knowledge in how to design a study using quantitative and qualitative methods, including the methods to use, how to combine the different methods to obtain a mixed method and what to consider when choosing a sample.

3.5.1 Qualitative and quantitative research

Qualitative and quantitative methods both have advantages and disadvantages when it comes to data collection. Qualitative methods are typically performed in order to gain deeper knowledge in questions opens for interpretation, for instance when the researcher addressing *why* a certain person react, feel or think in a specific manner(Given, 2008). There are various methods available for performing qualitative studies, where some of the most popular includes, interviews, observations, and diaries (McQuarrie, 2011). Qualitative studies are often used in an early phase to generate new theories and hypotheses, which later on the can be confirmed by quantitative data (Lobe, 2008).

Quantitative research aims to explain a phenomenon by collecting numerical data (Aliaga & Gunderson, 1999). Quantitative research is suitable to use for interpret questions like, *who, how many, and what* (Given, 2008). The data extracted is in the form of numbers and can therefore be quantified and summarized (Golafshani, 2003), furthermore, is the data appropriate to utilize in computerized mathematical and statistical tools (van Raan, 2013).

3.5.2 Mixed Methods Research

Mixed method research involves qualitative and quantitative collection of data in a study. This method has been increasingly popular during the last recent years (Creswell, 2013). One reason for this is according to (Sandelowski, 2000), the opportunity of expand research scope and improve the analytical strengths. Another reason is the complexity of the human phenomena that simply require more complex research methods.

Mixed research designs can either be fixed or emergent. Fixed design means that the use of both quantitative and qualitative data is predetermined from the beginning of the study, whereas the emergent design arises during the process of conducting a qualitative or quantitative study (Creswell, 2013).

The difficulty in using mixed methods is expressed in literature but in some researches there isn't enough to just use one of the methods (quantitative and qualitative) as one may be needed to provide complementing data to strengthen the data from the other. Creswell stress the importance for researchers to explicitly describe why mixed methods are used, since it is more complicated and not guarantee for a better result. (Sandelowski, 2000) also claim that there are still much confusion and relatively little direction when it comes to combine qualitative and quantitative methods.

Another important step when performing a mixed method study is to determine the level of interaction between qualitative and quantitative strands. (Greene, 2007) argued that there are two options available to choose, either independent or interactive. An independent level of interaction means that the researcher keeps the qualitative-and quantitative questions, data and analysis independent, only to mix them when overall conclusions and interpretation in the end of the study is performed. Keeping the study on an interactive level means a direct interaction between qualitative and quantitative strands during the study.

Creswell (2003) presents six different design strategies to consider when performing a study. “The convergent parallel design” is characterized by two or more methods used concurrently at the subject to confirm, cross-validate or corroborate data from a study. This design is useful when trying to overcome weaknesses from one method with strength from another. Another design mentioned is called the “Embedded design”, this strategy gives priority to one of the methods and provides guidance in the study whereas the second method is embedded. The purpose of this design is to be able to seek information from various levels of the subject (Caracelli & Greene, 1993). The resulting designs are somewhat sequencing qualitative and/or quantitative methods after each other depending on where the focus on the study is based on.

3.5.3 How to choose a sample

The most common method of quantitative sampling is to use probability samples, a random sample where all members has an equal chance to be selected (Marshall, 1996).

(Marshall, 1996) states that larger the sampling size, gives smaller risk of a random sampling error. But the author also points out that there is no gain in studying to large samples as the error is inversely proportional to the square root of the sample size.

The sample size in quantitative research can be calculated from four factors that need to be decided (Smith, 2013). First, how big is the total population size which your sample should represent? Secondly, how much error can you allow? No research is 100% error proof so you need to decide on a margin of error (confidence interval) to allow (e.g. +/- 5%). Thirdly, how confident do you want to be that you end up in between your predefined confidence interval? This is called the confident level (e.g. 95% confident). And fourthly, how big of a variance do you expect in your research? This is called standard of deviation.

These four factors can be combined and calculated with a formula providing the sample size needed to fill the requirements of quality in the research.

The sample size are often discussed in qualitative research as some may say that it’s non-important as the research is based on opinions and thoughts and doesn’t need to be validated by a large sample number. Other researches states numbers are important, (Sandelowski, 1995) states that there is no computation on how large of a sample is needed to offer validity to a research, but there’s a risk that an inadequate sample size can undermine the credibility of a research and a too large sample size doesn’t allow the deep analysis wanted in a qualitative research. (Ritchie, Lewis, Nicholls, & Ormston, 2013) states that the sample size of a qualitative study should stay under 50 for easier management and analyze, and if exceeded the researchers need to take into consideration how to maintain quality of research across the whole sample.(Ritchie et al., 2013) also states that it is important to not have a too small sample as it may miss key information and lack diversity within the sample.

The qualitative research is performed in order to gain a deeper understanding of a phenomena and to develop explanations (Ritchie et al., 2013). In order to gain this information the sample used needs to fulfill some criteria to be relevant in the research, and in comparison to quantitative research the sample can't be randomly picked. (Marshall, 1996) states that a random sample gives the best opportunity to generalize a population based on the results but doesn't support to develop a deeper understanding of an issue.

(Ritchie et al., 2013) suggests working with prescribed selection criteria for the sample. Some criteria may be interlocked to assure a certain number filling those criteria (e.g. age and gender to assure a certain number of men and women in all age categories) and then (Ritchie et al., 2013) suggests including a quota for each criteria to assure a good mix in the sample fitting the purpose of the research (e.g. 3-4 persons at age 13-17 years, 4-5 persons at age 18-25, etc.)

3.5.4 Qualitative interviews

The quantitative data in the project are to be collected with eye tracking technology but there is earlier discussion stating that the quantitative data collected may be needed to be complemented with qualitative data (Sandelowski, 2000).

Conducting personal interviews is a well-known approach for extracting qualitative data from subjects. (Rubin & Rubin, 2011) justifies qualitative interviews from the point that they let us understand what is often looked at but seldom expressed. (Weiss, 1995) express the importance of qualitative interviews for its ability to learn what people perceive and how they interpreted their perceptions. Furthermore, Weiss discuss the implications on conducting personal interview studies; the analysis will be much more dependent on interpretation instead of categorization and counting, due to the richness and subject unique data.

(Silverman, 2010) also discusses the advantages and limitations for qualitative interviews, and claims that the appropriateness of a method is highly dictated by the research question itself. Furthermore, he argues for the need of conducting qualitative interviews when the research question demands it and not be scared away of the fact that the research norm today implies quantitative methods. One reason for this might be the fact that qualitative data is less astatically oriented when displaying results.

(Katz, 1983) discusses some of the issues regarding field interviews. He claims that interview studies is especially vulnerable because of the researchers fail ratio when it comes to give; instructions, formulating questions, the order of the questions and in which situation probing is needed. (Foddy & Foddy, 1994) on the other hand argues that one of the key-issues when performing interviews is the understanding between the interviewer and the interviewee. The researcher needs to deliver the question in a way that the interviewee is able to decode the question and answer in a way that the interviewer expects the receiving message to be delivered.

In general there are three different kind of interview methods, structured, semi structured and un-structured. When using a structured interview method, the interview is limited by a script of fixed questions. The interviews are often conducted person to person or by telephone and can involve both open and closed questions (McQuarrie, 2011). According to (Wilson, 2013) structured interviews is most appropriate when the researchers are aware of a very precise problem and want to obtain detailed information about this specific matter. A structured interview requires in general less practice and cost less than semi structured and unstructured interviews. The main strength over semi and unstructured methods lies in the analysis of the data, since all the answers are based upon the same questions (Wilson, 2013).

A Semi structured interview is a combination between the structured interviews with room for open ended exploration, the interviewer sets up a general framework used in order to secure that the main questions are being asked with the same degree of freedom for all participants (Drever, 1995). Furthermore (Wilson, 2013) argues that it is vital to predefine probing suggestions and closing comments to make sure that the interview flows in a good manner. The major strength of conducting semi structured interviews lies in the ability to give the subject enough freedom to explain thoughts and underlying reasons about a specific question (Horton, Macve, & Struyven, 2004).

Asking the right questions in a semi structured interview, which lead the subject to the particular point of interest is of course vital. Even though this type of research is frequently practiced it is according to (Dunn, 2000) impossible to provide a strict guide for good practice when it comes to formulate interview questions. Further, he claims that every single case needs its own preparations and practice. (Flowerdew & Martin, 2005) agrees with this and claims that in social interaction there are simply no hard rules that can be followed. However, different types of questions can be categorized, (Harrell & Bradley, 2009) describes three different types of questions; descriptive, structural and contrast. The descriptive questions are in general asked in order for the subject to provide insights of areas that the researcher may not have thought of. Structural questions are used to understand a sets or list of things and how they relate to each other. One example of such question can be; why did you choose to buy a Volvo? Contrast questions helps the researcher to differentiate between the answers that already been obtained. One example of questions can be; why do you think you find this part disturbing by not this one?

3.5.5 Validity and reliability in research

The question of a study's reliability and validity answers whether the study studies what it is expected to study and this with the use of consistent measures (CSU, 2015). The validity and reliability of a study is the important core of what is accepted as scientific proof (Shuttleworth, 2008).

Validity

Validity involves the degree to which a performed study accurately reflects what the researcher is set out to measure. It can be further divided into external validity and internal validity:

- *External Validity* is closely related to generalization, and that the conclusions from a study can be generalized for a wider population at another time and space (Laerd-Dissertation, 2012). The results obtained in a study are solely based on the population used in that study, and external validity is extended to that the result from that smaller part of a population can represent the greater population which it is picked from (Shuttleworth, 2008).
- *Internal Validity* includes how well a study is performed and to the extent the researchers has "taken into account alternative explanations for any causal relationships they explore" (CSU, 2015). Can the researcher clearly state that the causal relationship observed is due to the variables studied and not affected by anything else (Boyd, 2015).

(Rogelberg, 2002) States that validity consists of four major features, which he bases on the definition of validity stated by (Messick, 1990). First it is to be seen as a judgment which is assessed and not measured. Secondly the judgment of validity is dependent on a combination of data and thoughts in order to be made. Thirdly it involves assessment of all the methods used, not just testing. The fourth feature requires more examination but arguing that one must investigate consequences of its conclusions. (Shepard, 1997) builds on that definition by highlighting the importance in looking at both the positive/negative and unintended/intended consequences of score-based conclusions.

Reliability

Reliability in research is based on the degree to which a method or procedure produces stable, trustworthy and consistent results. To assure reliability, the same study should be possible to be easily repeated by following researchers, leading to statistically significant results (Shuttleworth, 2008).

With ensuring stability in the research the reliability can be increased as "a high degree of stability indicates a high degree of reliability" (Golafshani, 2003). The stability of the measurement procedure is inclusive of many different variables, e.g. the measurement tools and the researcher performing the study, which introduces a degree of error into the study affecting the true score of the measurement (Laerd-Dissertation, 2012), and designing the study to minimize the risk of errors will increase the reliability of the end result.

The quantitative researcher measures with the use of an instrument constructed to be used according to predetermined procedures in a standardized manner. The question is, even if the procedure is well thought out and planned, does the measurement instrument measure what it is supposed to measure (Golafshani, 2003). If the reliability is dependent on the construction of the instrument in quantitative research, the researcher affects the reliability in qualitative research as "the researcher is the instrument" (Patton, 1990).

The importance of validity and reliability of a study

(Wainer & Braun, 2013) states that the researchers' involvement in quantitative research greatly reduces the validity of the test, as they affect the interplay between construct and the resulting data to validate their hypothesis. The concept of validity in qualitative research is discussed between researchers where the term validity often is substituted by other terms; one is the commonly used "trustworthiness" (Lincoln & Guba, 1985), another "quality" (Seale, 1999), and a third "rigor" (Davies & Dodd, 2002).

(Davies & Dodd, 2002) differentiates the use of rigor between qualitative and quantitative research, where the concept of bias in quantitative research is known and that rigor needs to be further analyzed through exploring of subjectivity and the social interaction in interviewing.

Some researcher's states that the concept of reliability in qualitative research is misleading due to that the purpose of a qualitative study is to generate understanding, in comparison to the purpose of explaining in quantitative studies (Stenbacka, 2001). The thoughts differ and there's also researchers stating that the reliability is important even in qualitative research, where some researchers redefine the term to be more specific. (Lincoln & Guba, 1985) uses the word "dependability" as a corresponding term to reliability in quantitative research, including the "inquiry audit" as a valuable measure. (Seale, 1999) highlights the importance of examining "trustworthiness" to gain reliability, and (Patton, 1990) states that "reliability is a consequence of the validity in a study".

Trustworthiness is a word brought up in literature when talking about both validity and reliability and (Lincoln & Guba, 1985) divides the terms into four questions that needs to be asked of research reports, their truth value, applicability, consistency and neutrality.

"If the issues of reliability, validity, trustworthiness, quality and rigor are meant differentiating a 'good' from 'bad' research then testing and increasing the reliability, validity, trustworthiness, quality and rigor will be important to the research in any paradigm."

– (Golafshani, 2003)

3.5.6 Triangulating

A combination of both qualitative and quantitative material is often used in order to complement each other. The reason for this is to increase the confidence of an outcome since both qualitative and quantitative methods have different advantages. Using the results from either a qualitative or quantitative study as a starting point for a new study is called triangulating. The use of triangulating can be traced back to 1959 when (Campbell & Fiske, 1998) argued that more than one method should be used in the validation process in order to ensure the validity from the results (Jick, 1979). Researchers utilizes different kinds of triangulations such as the "between (or across) method and "within method" labeled by (Denzin, 1978). The "between method" method is basically a cross validation when two or more distinct methods are showing comparable data, whereas the "within method" involves two or several techniques in a given quantified or qualified method used to collect and interpret data (Jick, 1979).

4. PRE STUDY

This chapter of the report will describe the method of work, display the results and discuss the results from the Pre Study with the focus of investigating the maturity of eye tracking technology provided by Ergoneers, to see to the possibility of using it as an evaluation tool of cars perceived quality. This is done by trying to find answers to the first research question, and subsequent questions.

RQ 1: *Is the Dikablis Essential eye tracking technology mature enough to be used as a tool for evaluation of a cars Perceived Quality?*

SRQ 1.1: *What are the problem areas affecting the validity and reliability of a conducted study?*

4.1 The methodology of performing the Pre Study

The Pre Study was divided into three parts each with an individual purpose.

1. Getting acquainted with the Dikablis Essential system
2. Evaluation of markers detection
3. Evaluation of the Calibration Procedure

The reason to the dividing of the Pre Study were due to that two bigger problem areas were identified in the first step, the markers and the calibration procedure, and it was decided to conduct two studies solely focusing on these two areas.

4.1.1 Preparation for the Pre Study

First an introducing demonstration of the Dikablis system and the analyze software D-lab 2.5 was given at the HMI-lab at Volvo Cars. The demonstration was held by Volvo Cars employees with previous experience of the system. Complementing reading was done in the Dikablis Essential Handbook (Ergoneers.Group, 2011) and the D-lab 2.5 Manual (Ergoneers.Group, 2012).

During the Pre Study a Volvo V40 was used as a mediating tool to help get a better understanding of how the eye tracking system could be used in collaboration with a vehicle (see Fig.17).

Figure 17: A Volvo V40, the model that served as a mediating tool during the Pre Studies(Paultan.org, 2013).

The V40 XC was parked in a venue called the "Analyze Lab" at Volvo Cars, a clearly lit area offering enough space around the car to allow for easy movement.

An associate at Volvo Cars assisted in the study in practice to calibrate the system and to collect data for analyze.

4.1.2 Getting acquainted with Dikablis Essential

The first study was conducted with the objective of getting acquainted with the eye tracking system including the different functions and features.

The associate was seated in the front seat of the V40 and was equipped with the glasses. Several markers in different sizes were attached at clearly visible places both in the interior and at the exterior of the car, an example of interior placement can be seen in figure 18. The amount of markers was in this stage excessive in number to investigate many different potential placements at once.

A calibration was made to adjust the system to the associates' eye, and then the associate was asked to visually evaluate the interior while staying positioned in the driver's seat. There were no directions or guidance in how the associate should behave as the evaluation was totally free to perform.

After the eye tracking data was recorded of the interior, the exterior was also evaluated whiles the associate moved around the vehicle. The eye tracking recordings were later imported in D-lab 2.5 where it was observed to look for eventual problem areas together with practice in analyzing the data with the different built in analyze tools.

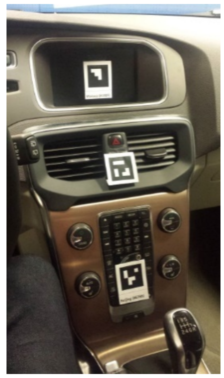


Figure 18: Example of attachment of markers on the Instrument Panel and center Display.

The procedure of recording data and analyzing it where repeated several times in order to gain experience in working with Dikablis Essential.

During the repetitive use of the system the calibration procedure and analyze were continuously practiced, and the markers were varied in size, placement and total number, together with continuous documentation on potential problems to be investigated further.

4.1.3 Evaluation of Markers Detection

From the first part of the Pre Study the markers showed to have some limitations and it was suspected that some markers were easier recognized by the eye tracking system than others. The limitations led to the need for further testing of the markers and an evaluation and comparison between the markers was performed see the difference in recognition between them.

All markers were used in the evaluation and printed in two different sizes resulting in 32 markers which were attached on a white wall in good lighting conditions. 16 markers in the size of 10x10 and 16 markers of size 15x15 were attached side by side in an alphabetical order (see Fig.19).

The rotation of the markers was not taken into consideration as Ergoneers Group has stated in the manual that the rotation of a marker doesn't affect the recognition of it (Ergoneers.Group, 2011).

Figure 19: Display of how the markers were attached during the evaluation of recognition rate

The size 10x10 and 15x15 markers were looked at separately as a marker can only be recognized at one place at a time. Meaning that in one recording each marker can only be used ones, leading to a maximum amount of markers in a recording to 16.

The eye tracking system was attached and calibrated on one of the test leaders which started by looking at the wall from a distance of 4.5 meters, moving closer by 0.5 meters intervals, ending up at the distance of 1 meter from the wall. At each distance the wall was looked at approximately 10 seconds.

The video was afterwards analyzed with D-lab to see when the markers were recognized and to examine the individual performance of each marker. Five video clips (one at each distance), with the length of five seconds, was cut from the entire video and analyzed in shorter frames.

Since the field-cam observing the environment at 25 HZ, every frame for analysis was 0.040 seconds resulting in a total of 125 picture frames for every five second clip. The evaluation between markers were done by analyzing how many frames out of the total 125 they were recognized and to compare the percentage in attention to establish a ranking order of recognition between the markers.

4.1.4 Evaluation of the Calibration Procedure

The first part of the Pre Study did also show on limitations in the sensitive procedure of calibrating the system and that the calibration was to be performed with high precision in order to get good validity of the recordings. An evaluation of the calibration procedure was made in order to gain better knowledge of parameters affecting the calibration.

The evaluation was done by one test leader attaching the system to the other test leaders' head and then performed ten repetitions of calibration using two different approaches. The first five repetitions were conducted using a wide square of placement of the calibration points, whereas the last five repetitions used a narrower square (see Fig.20).

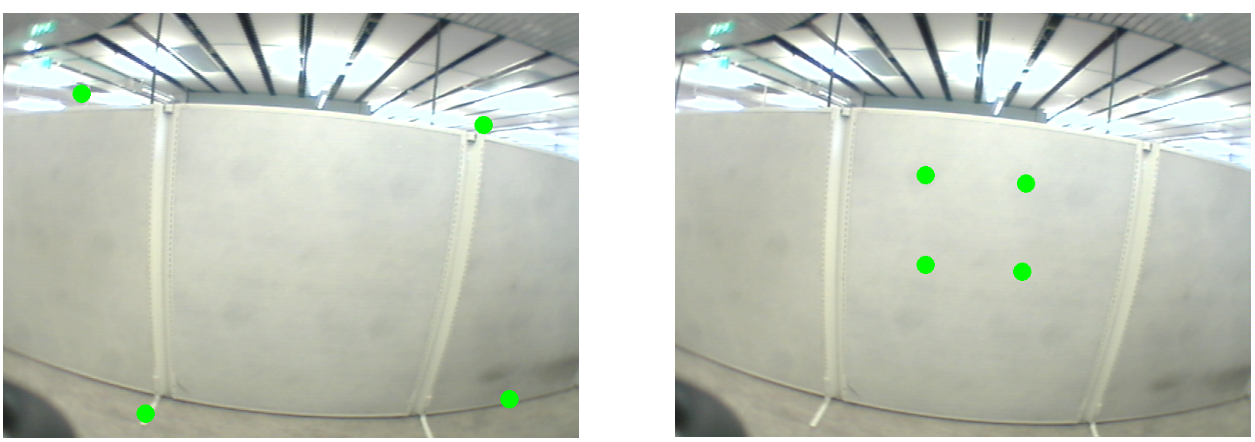


Figure 20: To the left, a wider placement. To the right, a more narrow placement (Print screen from Dikablis)

In order to evaluate the performances from the different approaches, recordings were made where the test leader equipped with the glasses were asked to gaze at a moving marker following a pre-defined pattern at constant speed. Performing the recordings as similar as possible aimed at giving validity to the evaluation. The design of this test where deliberately designed to push the limits of the system, this was decided in order to easier differentiate which type of calibration that supported the highest recognitions rates.

The data received from the study was then analyzed using D-Lab 2.5. Three different sizes of AOI's were drawn at the "project" of the calibration test to assure that the same AOI's were used for calculations of all the recordings (see Fig.21). The glance metrics used to analyze the calibration was; Total Glance time (GTOT), Number of Glances (GNUM), and Attention Ratio (ARATIO). The Recognition Rate (RR) of each recording was also documented to establish the rate of the pupil detection.

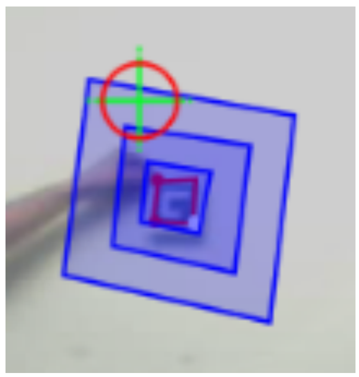


Figure 21: Three AOI´s defined to one single marker (Print screen from D-lab 2.5)

4.2 The result from the Pre Study

This chapter will report the results from the Pre Study including evaluation of marker recognition and the calibration procedure.

4.2.1 Markers

The Pre Study showed on delimitations affecting the recognition of the markers, which will prevent creation of AOI's and complicate analyze.

There were issues shown in the Pre Study that prevented recognition of the markers, and one issue was covering the marker while interacting with the car, interactions such as using buttons and controls or to feel materials. When the hand covered a part of the marker, it was not recognized and the AOI was not visible (see Fig.22).

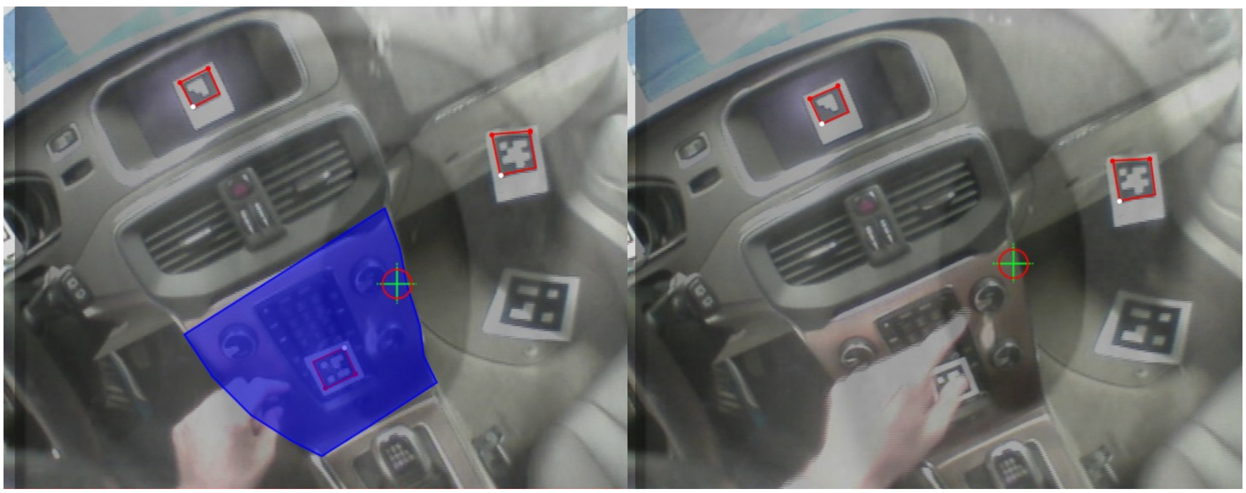


Figure 22: To the left, the marker connected to the AOI of the IP was visual. To the right, the marker was covered while using the buttons on the IP and the AOI was lost (Print screen from D-lab 2.5).

Another issue that affected the recognition was the markers sensitivity to reflection of light. If the markers directed towards a stronger light source, they reflected the light, which resulted in that the white areas of the marker melted together and that the marker couldn't be recognized (see Fig.23).

During the exterior testing it was shown that placing the markers at the inside of the car, underneath the glass, protected against reflection (see Fig.23). But the reflection in the glass still disturbed the recognition when it ended up at the marker during movement around the car (see Fig.23).

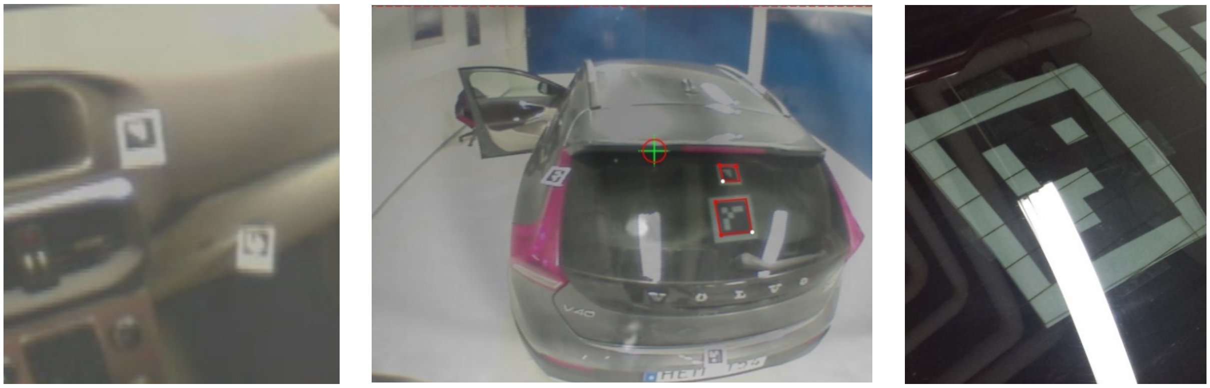


Figure 23: To the left, white area of markers has melt together. In the middle, the markers could be recognized when hidden under the car window. To the right, the marker is not detected due to a reflective light source in the glass.

The creation of markers using Smart Track showed to be too sensitive when moving around. There were no objects with high enough contrast to offer a stable enough marker as a supplement to the ones provided with the system.

In addition to the issues affecting recognition it was shown that the markers were recognized differently good by the system and all the markers were ranked in a list from 1-16 out of mean recognition rate (see appendix 2). It should be notices that these recognitions rates where obtained during the tests using the longest distance between the eye tracking and the markers. This distance is used in order to clearly differentiate between the marker performances. Hence the rather low recognition rates where obtained. Using a distance of 1m all of the markers where recognized over 95 percent of the time.

The result showed that one marker where better recognized than all the rest. This marker was Monaco , with a mean recognition rate of 70,25 percent. Following Monaco in top five was Sydney (60,80 %), Santiago (59,95 %), Dublin (56,6 %), and Seoul (56,5 %) (see Fig.24).

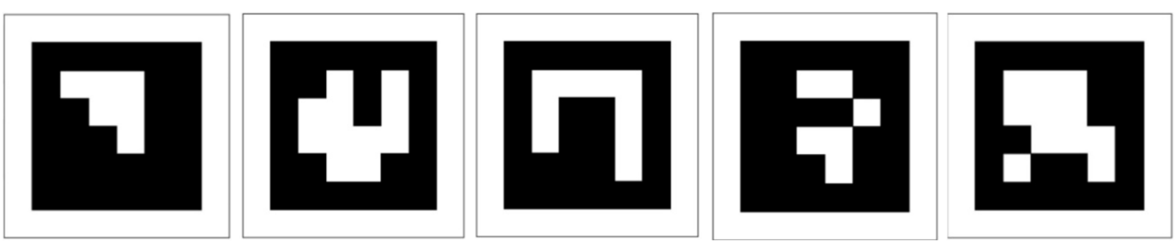


Figure 24: The markers placed in the top 5, from left to right: Monaco, Sydney, Santiago, Dublin, and Seoul.

The worst recognized markers were Washington, placed 16th with 38,25 percent, just beaten by Ankara, placed 15th with 38,7 percent (see Fig.25).

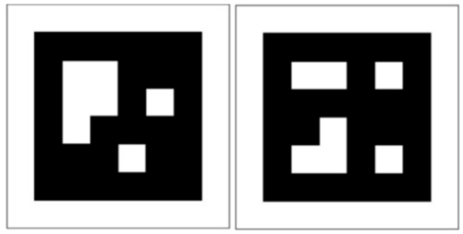


Figure 25: The worst markers in recognition, from left to right: Washington, and Ankara.

4.2.2 Calibration Procedure

Performing the calibration evaluation showed on three different issues. The first issue was the difficulties in keeping the head and neck static when calibrating. The second issue was the sensitivity in achieving the optimum settings of adjusting the eye- and field camera manually. The third issue was that it was shown that the detection of the pupil was highly affected by the anatomy of the person wearing the glasses (e.g. the wide of the eye opening, and the length of the eye-lashes).

The results from the evaluation of the differences between wide and narrow placement of the calibration points showed on significant differences in performance. The approach with shorter distances between the points was increasing the systems performance with higher mean attention ratio (ARATIO) in all of the interesting subjects (see Table 3).

The narrow calibration procedure received an attention rate at the AOI size 1 of 20,247 percent, while the wider points only received 2,8 percent. The difference in attention rate was big for the bigger AOI's as well, but the difference in number of glances wasn't as significant. The wider placement did even get a higher number of glances when evaluating the AOI size 3 with 29,2 glances compared to the 24,5 glances in narrow placement.

Table 3: **The measurements of narrow and wider placement of calibration points**

Calibration Procedure	Mean Total glance time (Sec)	Mean Number of Glances (No)	Mean Attention Ratio (%)
Narrow Placement			
AOI Size 1	6,206	14,6	20,247
AOI Size 2	9,729	18,2	31,254
AOI Size 3	16,212	24,5	49,202
Wider Placement			
AOI Size 1	1,008	7,0	2,826
AOI Size 2	1,720	11,8	4,727
AOI Size 3	9,992	29,2	26,638

The full table of calculated glance metrics from the study can be found in Appendix 3.

4.2.3 Areas of Interests

It was shown when drawing AOI's in D-lab that the stability of the AOI's was highly dependent on the placement of the controlling marker. Every AOI needed to be connected to a marker and the longer the distance from the marker to the AOI, the more unstable was the AOI in keeping its correct position during the analyze. Figure 26 shows the placement of the AOI at the display when controlled by the marker centered at the display, while figure 27 shows the placement of the AOI while controlled by the marker at the steering wheel. The two set-ups of pictures were taken at the exact same picture frame in the recording.

Figure 26: Picture of a stable front display AOI when directed by the centered marker on the dimmer (Print screen from D-lab 2.5).

Figure 27: Picture of a more unstable front display AOI when directed by the marker on the steering wheel (Print screen from D-lab 2.5).

4.3 The discussion about the Pre Study

Eye tracking has been proven to be an adequate and valuable evaluation method for other market research projects (see chapter 3.3.3). While the earlier projects referenced have been performed in a 2D environment and with the use of eye tracking systems with different developer, the discussion will focus more on how it would work in a 3D environment and solely focusing on the eye tracking system Dikablis Essential available at Volvo Cars.

In order of trying to find answers to the research question each potential problem area will be discussed individually. The problem areas are the *Markers*, the *Calibration Procedure*, the *Analyze of Data*, and the *Sensitivity of the System*.

4.3.1 Markers

Using the markers is to be seen as one of the most sensitive part in order to obtain statistically proven data from the eye tracking study, as they need to be recognized by the eye tracking glasses at certain desired moments of the recording in order to later assist in the creation of AOI's in D-lab 2.5.

To assure high recognition of the markers and to reduce the risk of them being covered it is suggested to place the markers where no interactions are made with the car. It is also suggested to use the markers which performed best in the evaluation of markers as in large extent as possible.

If the markers are exposed to reflection it increases the risk that the white areas in the markers are floating together reducing the recognition of the marker. It is suggested to place markers to be directed away from light sources, and to investigate a possible method of making the markers more matte. In the exterior evaluation it is suggested to place the markers beneath the windows to shade them, and to use more than one marker for an area to assure that one is always visible for the field camera in the case of an direct reflection of the light source ends up over an marker (see Fig.28).

Figure 28: Picture showing how one marker is still visible when one is covered by a reflecting light source.

It is also important to place the markers in a way that they are still visible for the field camera when the person equipped with the glasses is moving around. It was shown that if the angle between the field cameras view and the markers is too sharp the markers wouldn't be recognized by the field camera (see Fig.29).

Figure 29: Picture showing that too sharp angles are easily obtained in exterior evaluation (Print screen from D-lab 2.5).

It is thought that the markers draw a lot of attention from the actual evaluation of the car and some suggestions is stated in order to reduce the risk of the markers to draw attention.

- It is suggested to keep the size of the markers to a minimum, but still large enough to be detected by the field camera.
- It is suggested to keep the amount of markers to a minimum, but still enough to cover all the desired AOI's to be used in the final Main Study.
- It is suggested to place the markers at a distance from the AOI which it controls and don't have any markers placed within an AOI to be used in the final Main Study.

In addition to the lessons learned from the Pre Study it is suggested to place the markers according to the guidelines provided by Ergoneers Group (Ergoneers.Group, 2012).

4.3.2 Calibration Procedure

The calibration procedure is also to be seen as a sensitive part in the work of using an eye tracking system. The performance of the calibration procedure is greatly affected by the experience and knowledge of the person performing the calibration, and it is suggested to continue practicing the calibration procedure to increase precision before performing the final Main Study.

From the Pre Study some areas of improvements were found to ease for an increasing precision of the calibration.

The clearest area of improvement is to use a narrower placement of the calibration points during the calibration, and it is suggested to have portable calibration points that can be adjusted according to each and every person. One answer to the increase in performance for the narrower placement may be that using a wider calibration requires glancing at the periphery, which in turn reduces the focus possible to keep on a specific point, and therefore also reduces the precision in the calibration.

Another interesting finding which adds proof to that the narrower placement is preferred is that the total number of glances between the two procedures of calibration is similar in number but the narrower has a much higher attention ratio. This would mean that the narrow approach is more stable than the wider as the wider calibration videos includes a glance pattern that passes over the AOI's more than stabilizes on them.

It is also suggested to place the calibration points at a distance as similar to the distance of the objects visually evaluated in the actual study, and to construct the calibration points to contrast with the background, to assure high focus and acuity.

The last suggestion to assure high performance of the calibration is to help the person wearing the glasses to focus in holding their head still by offering a good seating arrangement with a stable head rest.

4.3.3 Analyze of Data

The last area is analysis of the data collected with the eye tracking system. The eye tracking data can be imported into D-lab 2.5 in order to calculate glance metrics showing quantitative data on where a person is looking the most and first, but there isn't any information on why a person is looking at an area more than others.

In order to find answers on why, there is need to include a qualitative data collection method into the project, and it is suggested that the method is built on interaction with the participants of the study to obtain information.

4.3.3.1 Analyze in D-lab 2.5

Analyze in D-lab 2.5 is directly dependent on the placement and recognition of the markers to be used in drawing AOI's. A marker is chosen to control an AOI, and when the marker is recognized then the AOI is visible and able to collect glance data. The distance between the controlling marker and the AOI to be controlled affects the stability of the AOI.

To offer stability to the AOI, it is suggested that the AOI should be drawn when staring directly on the desired AOI (see Fig.29). It is when the gaze is within the AOI that it is important that the AOI is stable. When the gaze isn't within an AOI, the collection of data isn't needed, and if the AOI is unstable it doesn't matter as long as it stabilizes again when the gaze is within the AOI.

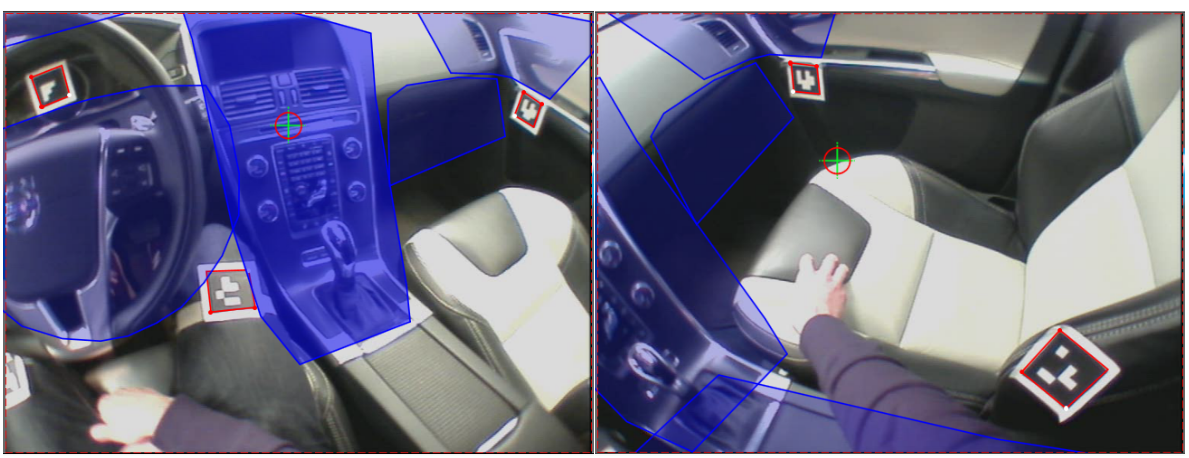


Figure 29: To the left, AOI drawn while direct staring at the center stack. To the right, instability of the center stack as the AOI has been drawn at the wrong moment (Print screen from D-lab 2.5).

As the distance between marker and AOI affects stability, it is suggested that AOI's should be controlled by markers closely positioned to the desired AOI. This will be in conflict with the suggestion of not placing markers within AOI's as they may disturb and attract attention. It is suggested to perform further investigation of how to place the markers to find the optimal length between the markers and the AOI's.

There's a risk of reducing validity if the AOI's are drawn for each person participating in the study as there may be a difference in shape and size of the AOI's covering the same area. It is suggested to draw the decided AOI's as "global AOI's" at the highest level in D-lab 2.5, "project", covering all underlying "subjects" of that study to assure that the size and shape of the AOI's are the same throughout the analyze.

In order to ease for analyze it is suggested to record a Reference Recording. A recorded eye tracking video made by the test leaders which includes "perfect" views of each AOI to be used for all recordings. The "perfect" AOI's are created when the test leader stares at each chosen area, to easier draw good AOI's in D-lab 2.5.

How small, detailed and specific the AOI´s can be drawn is very dependent on the performance of the calibration. Even though the calibration procedure can be practiced to increase the accuracy, it is suggested to keep the AOI's bigger and to focus on bigger areas including several components instead of component by component (see Fig.30).

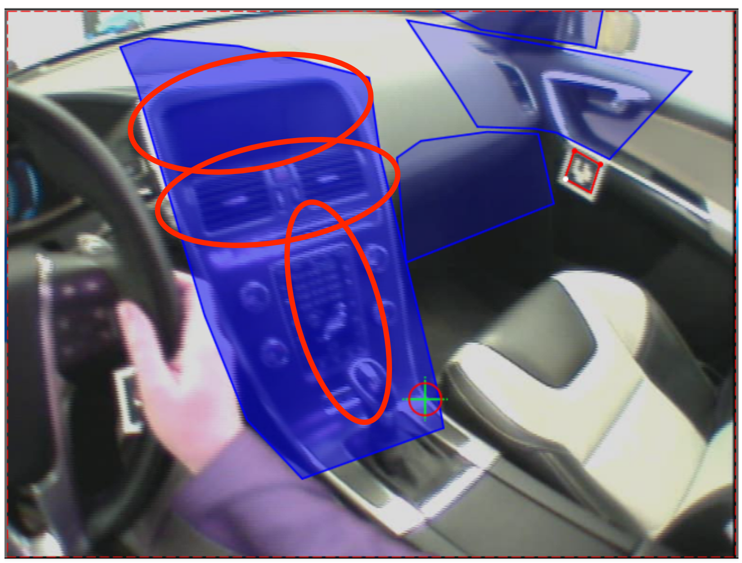


Figure 30: A potential AOI, the Center Stack, including smaller components such as the stereo, central air vents, and the central display (Print screen from D-lab 2.5).

4.3.4 Sensitivity of the System

From observations of the eye tracking recordings it was shown that the pupil detection was very sensitive and a reason may be the eye movements used when evaluating the car.

In order to calculate glance metrics in the recordings it is needed for glance fixations within an AOI. The glance fixations need to be long enough to be considered to evaluate the AOI and not just only crossing through.

Even though there weren't more than three different persons recorded in the Pre Study, the results from observations showed on different eye movement patterns between the persons. Some movements were quicker with extensive saccadic movements repeated at several times per second while other recordings indicated smooth pursuit movements.

It was clear that the eye tracking system had more problems catching the saccadic quick movements, and it is thought that during the Main study there will probably be more persons using quicker eye movements when evaluating the car. This may be a problem, as some data may be lost in comparison to the ones using smooth pursuit in evaluation.

Another potential problem is when the person wearing the glasses only moves their eyes and not the head. This type of eye movement was common when evaluating areas low or high placed in the car (see Fig.31). It is therefore suggested to always encourage the person using the eye tracking glasses to follow around with the entire head when evaluating the car.

Figure 31: To the left, using the entire head to look at an area low placed. To the right, only using the eyes to look at an area low placed.

4.4 Answer to RQ1 and Conclusions

This Pre Study was performed in order to answer RQ1:

RQ 1: *Is the Dikablis Essential eye tracking technology mature enough to be used as a tool for evaluation of a cars Perceived Quality?*

SRQ 1.1: *What are the problem areas affecting the validity and reliability of a conducted study?*

No explicit threshold values where established before the Pre Study, which would have objectively answer if the system were mature enough, or not. Instead a subjective assessment of the systems performance in total where used to answer RQ1.

The obtained results from the Pre Study showed that the Dikablis Essential eye tracking system is mature enough to conduct perceived quality evaluations. There is however several of issues that needs to be considered in order to gather valid data. These issues have been identified and also serve as answer to SRQ1.1. Hereinafter follows a condensed bullet list of conclusions made during the study that needs to be achieved and considered if answering Yes to RQ1.

Markers

- It is important to follow the guidelines in marker placement provided by Ergoneers Group (see chapter??).
- To reduce the risk of markers drawing attention there should be no markers placed within an AOI used in analyze.
- To reduce the risk of markers drawing attention the amount of markers, and the size of markers should be kept to a minimum.

Calibration Procedure

- It is decided to use a narrow placement of the calibration points as it is proven to give higher precision in calibration than a wider placement.
- The distance to the calibration points in calibration should be equal to the distance to the car when evaluating it with the eye tracking system to offer best focus.
- In order to provide stability to the person to be calibrated to obtain high precision in calibration it is decided to focus on given good seating arrangements with good neck stability.

Analyze of Data

- There is a need for a qualitative data collection to answer the question on “Why?” a person’s looks at an area more than other areas.

- The qualitative data collection method should include an interaction with the participants of the study to better control how the answer on the question “Why?” can be obtained.
- The AOI’s drawn in analyze should be the same for all participants to provide better validity to the result.
- A “Reference Recording” should be recorded to assist in drawing the AOI’s to work for all participants.
- The AOI’s needs to be bigger areas including more than one component as the systems precision aren’t meeting the requirement to analyze smaller areas.

Sensitivity of the System

- Encourage the participants to move their head and not only visually evaluate with their eyes to get better recognition rate of the pupil.

5. PILOT STUDY

This chapter of the report will describe the method of work, display the results, and discuss the results of the customer oriented Pilot Study with the focus on investigating the possibility to construct a customer oriented study using eye tracking technology. This will be done by trying to find answers to the second research question, and subsequent questions.

RQ 2: *How is a customer oriented study of Perceived Quality of cars using eye tracking technology constructed?*

> **SRQ 2.1:** *What are the problem areas affecting the validity and reliability of a conducted study?*
>
> **SRQ 2.2:** *Which qualitative data collection method should be used to increase the understanding of the eye tracking data?*

5.1 The methodology of performing the Pilot Study

The Pilot Study was structured to use the knowledge gained from the Pre Study and to involve people simulating potential customers. The aim of the Pilot Study was to be able to investigate how a customer oriented study could be constructed.

5.1.1 The structure and procedure of the study

Five associates at Volvo Cars assisted in the Pilot Study. A Volvo V40 was ones again used as a mediating tool and was placed in the Analyze-lab.

Markers were placed at the interior and the exterior of car, and in the Pilot study was also three chosen DLO areas included (the trunk, the driver's door, and the back right passenger door). The markers used were minimized in number and size because of the thought that they distract, and placed according to the suggestions stated in the Pre Study. Which markers to use were chosen according to the results from the evaluation of markers in order to assure that the best possible markers were used.

When attaching the markers, the test leaders also considered the placement of markers in accordance to give good analyze in D-lab, meaning that the markers should be placed near AOI's but not within any AOI.

To provide explanatory qualitative information in addition to the eye tracking data, it was decided to use the method qualitative interview and to perform one interview with each participant. A template for the interviews was compiled to assure that the same questions were asked to all participants.

The five associates were one by one invited to the Analyze-lab. When they arrived they were introduced to the study and told that they would evaluate the quality of the car. The persons were equipped with the eye tracking glasses and seated in a chair with good neck stability to ease for calibration.

The chair was placed at a distance of 1,5 meters from a white wall, representing a distance often used when evaluating the exterior of the car. The calibrations for the interior evaluation were made on the persons while they were seated in the driver's seat.

On the white wall in front of the chair were five points attached to be used in calibration, one to offer focus when looking straight ahead and four points to serve as calibration points. The calibration points were possible to move to get an adjustable narrow placement suited for each participant.

When a person had performed two visual evaluations, one of the interior seated in the driver's seat, and one of the exterior moving around the car, he/she were asked some questions. The questions included complementing facts about their perception of the car, how they experienced the execution of the study and their relationship with cars.

The collected data, both quantitative and qualitative, were analyzed. The analysis of quantitative data was made with calculations of glance metrics in D-lab 2.5 and the qualitative data through analyze of the answers from the interview questions.

Four different glance metrics were calculated (*Total Glance Time*, *Number of Glances*, *Mean Glance Duration* and *Attention Ratio*) to give an idea of the amount of data the evaluation would result in.

The quantitative calculations were mainly used to give extra information on which AOI's to include in the Main study. This was made by looking to the AOI's which was resulted to be most looked at together with the two test leaders own experiences to establish the final AOI's chosen for the Main Study.

The ARATIO of all AOI's in both exterior and interior evaluation were summarized to show the total percentage of attention within any of the AOI's to see where the persons looked at the most.

The qualitative analyze were made through observations of the eye tracking recordings in D-lab, looking to the answers from the interviews, together with notes taken during the study.

5.1.2 Additional methods

Four other issues were evaluated to see to potential problem areas in addition to the structure and procedure of the study.

Making the markers more matte

First issue was to make the markers more matte in order to reduce the effects of reflection, and three different methods were tested.

First method was to spray the printed markers with a matte spray, which should make the surface less reflective and more repellent to dirt. Second method was to try creating the markers by cutting and pasting matte black and white paper in the same shape. Third and last method was to evaluate the possibility to laminate the markers with a matte plastic film.

The newly created markers were analyzed with the eye tracking system, making a recording together with a normally printed marker to see if there were any improvements in reduction of reflections.

Estimating a time plan for the study

The second issue to evaluate is the total time it would take to perform the Main study and to construct a time plan.

Time for each task was documented through the pilot study and these times were later used in estimation of the time needed and used to create a final time plan for the main study. The biggest estimated time consuming actions in the Main study was *Calibration*, *Recording of Eye tracking Data* and the *Interviews*. These three areas were evaluated to give and idea of how long time each task would take.

The time of recording of eye tracking data is estimated by calculating a mean value from the times of each individual recording. The total time for a recorded eye tracking video was named Time for Evaluation (TE). The Time for calibration and interview was estimated.

The total time for one person was estimated by summarizing all the values, see Formula 1.

$$\textit{Total Time} = \textit{Calibration} + \textit{Interior Eval.} + \textit{Calibration} + \textit{Exterior Eval.} + \textit{Interview}$$

Formula 1: The formula of all factors which adds up to the total time for each person

Planning on which sample, location and car to use

The third issue is the preparations regarding which sample, location and car to use for the main study were held. Different stakeholders were consulted to get different inputs on pros and cons for different alternatives.

Evaluating the effects from the restricted movability

The restricted movability due to the great amount of cables following with the system was evaluated to see the effects it has on the participants' movement and experience of the study. This was made both by analyzing the recordings, observing their movement when evaluating the car, and to ask the persons about their experience after the evaluations.

The first three participants were assisted by the authors in carrying and handling the cables to ease for movement. The two last persons were equipped with a backpack where most of the cables and equipment were put to make it easier for them to move. All the cables which were needed to be carried can be seen in figure 32, with one cable inserted into the computer, one leading to the glasses and one inserted into an electric outlet.

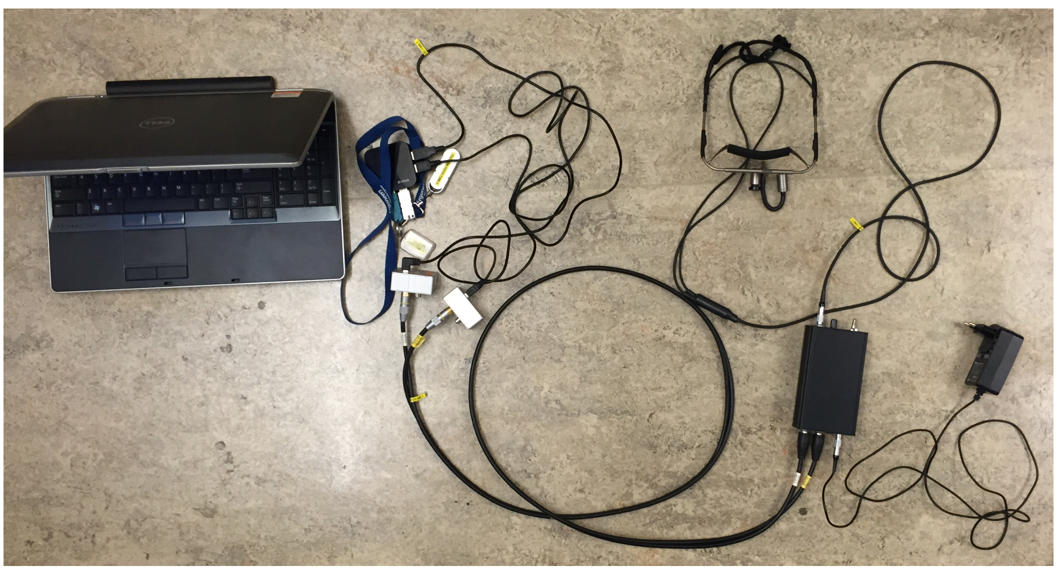


Figure 32: All the cables affecting the movability

5.2 The result from the Pilot Study

This chapter will present the quantitative and qualitative results from the customer oriented Pilot Study.

5.2.1 Quantitative data

The results of the quantitative data calculated through glance metrics in D-lab showed that in the interior evaluation of the car, only 22,85% of the total fixations were within any of the AOI's. There are two explanations for this low rate. Firstly, during analyze of the data is the system only evaluating fixations towards an area. This means that when the participants are seeking for new areas to evaluate and the gaze vector jumps from area to area it will not be included as a fixation. The second reason is that all of the areas inside the interior are not covered by predefined AOI´s.

The calculations of glance metrics did result in a lot of data possible to use to see the glances length in time, amount of glances, and in which AOI the attention was the highest. The total table of the glance metrics calculated can be found in appendix 4.

5.2.2 Qualitative interview

The interviews conducted after the visual evaluation gave complementing answers on what the participants considered important when evaluating a car and what they mainly looked at in evaluation.

All of the five participants answered that the first focus when evaluating the car was to see the overall design. They want to start by looking obliquely from the front to see proportions and relations between bigger components, such as the headlamps and the rims. Four out of five clearly stated that the functions in the interior were important and that they looked at buttons and controls. Evaluating where things were located and how to perform functions. One of the participants summarized it as:

"I pretend to drive away, and evaluate all the functions needed to do so".

- Test Person 3

5.2.3 Observations

Through observations of the eye tracking recording it was seen that attention was put on some areas of interest not included in the study. One area showed to be of interest for all participants were the backseat with estimated durations of looks between 3 and 20 seconds (Mean value 10,2 seconds). Other areas which were shown interest were the glove compartment and the sunshades.

All of the participants did interact with the car physically by simulating to drive, and throughout the sample it was common to adjust the rear view mirror, the seat, push buttons and open compartments. These interactions with the car did in some cases cover the markers and removed AOI's.

Another potential problem seen through observations was that some AOI's disappeared in the recordings at different times. Reason could be due to bad angle or high reflection of the markers, but also that the markers disappeared out of the field-cameras view in some situations (e.g. when moving closer to the car to investigate a certain component/part) (see Fig.33).

Figure 33: To the left, markers are clearly visible. To the right, marker disappears out of sight when to close to the car in exterior evaluation (Print screen from D-lab 2.5).

5.2.4 Making the markers more matte

In the observations of the eye tracking data it was clearly shown that one of the methods of reducing reflections was better than the other. In making the markers by cutting out matte black and white paper gave very good results while the spray and the lamination didn't give enough improvement to increase analyze.

5.2.5 Estimating a Time Plan

To construct an overall time plan for the Main Study was needed to offer data to estimate how long the actual study would take. The main tasks to perform during the Main Study are Calibration, Recording of eye tracking data, and a Qualitative interview. The estimated time for Calibration was stated to be 5 minutes and the estimated time for performing the interview was stated to be 15 minutes.

The third main task was calculated through Time for Evaluation (TE) which were between 122,52 – 274,92 sec in exterior evaluation and between 98,68 – 220,68 sec in interior evaluation (see table 4). This gave a mean value of 149,10 seconds on the interior evaluation and 181,88 seconds on the exterior evaluation.

Table 4: **Table presenting the Time for Evaluation (TE)**

TP	Interior TE (sec)	Exterior TE (sec)
1	123,60	153,52
2	120,80	274,92
3	181,76	207,32
4	98,68	122,52
5	220,68	151,16
Mean	149,10	181,88

Summarizing the time of the different parameters resulted in a mean total time of 30,5 minutes for each participant (see Formula 2).

$$\textbf{\textit{Total Time}} = \textbf{\textit{Calibration}} + \textbf{\textit{Interior Eval.}} + \textbf{\textit{Calibration}} + \textbf{\textit{Exterior Eval.}} + \textbf{\textit{Interview}}$$

$$\textbf{\textit{Total Time}} = \mathbf{5 + 2,5 + 5 + 3 + 15 + 30,5}\ \textbf{\textit{minutes}}$$

Formula 2: The calculations of the mean total time for each participant

5.2.6 Sample, location and car to use

The sample for the Main Study was frequently discussed in the project. Due to the time-frame and budget of the project it became evident that a trade-off between external validity and time/resources were needed. Four different scenarios were chosen as potential methods to attract test-persons and to perform the study. These four methods showed on different advantages and disadvantages (see table 5).

Table 5: **Summarizing table over the four different methods of sample collection**

Scenario	Validity	Cost	Ways to attract TP	Difficulties to attract TP	Venue for the study
Purchase sample from external company	High	High	Indirectly paid by Volvo Cars to perform the study	Low	Volvo
Students at Chalmers	Low	Medium	Cash or curiosity of new products (XC90)	High	Chalmers
Volvo Ergonomics Pool	Medium	Low	Company goodwill	Medium	Volvo
Mix of students and regular people	Medium	Medium	Cash or curiosity of new products (XC90)	High	Volvo, Public

5.2.7 Restricted movability

That the restricted movability was a big problem was clearly shown trough observations of the participants when evaluating the car as well as when looking through the eye tracking recordings. It was clear that during the exterior evaluation the participants moved unnaturally with slow and jerky movements. It was also something clearly mentioned during the interviews, with one person expressing;

"I would have moved a lot faster if I could. I want to get that overall perception first and may have gotten stuck on details in this low pace as I usually wouldn't care about".

- Test Person 1

5.3 The discussion about the Pilot Study

Involving potential customers in the work of using eye tracking technology gave more input on how the work needs to be structured and organized. The discussion includes differences to take into consideration in evaluating the interior and the exterior of the car.

5.3.1 Quantitative results

The result on the quantitative data showed on a low percentage of focus within the AOI's, this even though the focus has shifted from components into larger areas. A possible answer to the results is that people tend to move around a lot with their glance when evaluating a car.

Even though that the important AOI's in exterior evaluation is covered, the percentage of focus is beneath 20 %. Through observation it shows that the persons looks at the AOI's which is included in study but that the AOI's drawn in analyze is too unstable to catch all glances.

It was during the discussions of the Pilot Study it was stated that it would be beneficial to be able to establish where a person looks at first using the eye tracking technology. It is suggested to include this topic in the Main Study.

Looking through the result of the entire eye tracking data calculated, discussion started on which metrics important in order to answer the research questions. It was suggested to minimize the amount of data by reducing the number of glance metrics to calculate and the suggestion for the Main Study is to use Attention Ratio (ARATIO) to analyze where a person's looks at the most and Time to First Glance (TFG) to analyze where a person looks at the first. The other glance metrics were seen as superfluous.

5.3.2 Qualitative results

To include the qualitative method of data collection with interviews proved to be very beneficial in providing good complementary information to the eye tracking data.

It was shown that there is a possibility for different persons to interpret the questions differently and it is suggested to create a clear and structured guide on how to move through the study and how to ask the questions during the interview.

A result which can be further evaluated is the focus different persons had when evaluating the quality. During the Pilot Study it clearly showed to be more focused on functionality, but it is believed to be more categories that can identify different persons when it comes to evaluating the quality of a car.

The qualitative interview constructed in the Pilot Study needed to be improved, and it would be beneficial to gain knowledge from more experienced researchers within the area. It was clear that focusing more on qualitative collection of data will demand higher knowledge in interviewing techniques, and the stating of questions are crucial in arriving at the right information.

5.3.3 Interior evaluation

The result showing that the precision increases if the distance during calibration is the same as during evaluation led to a suggestion to perform the interior calibration while seated at the driver's seat. A disadvantage is that the information on where people tend to look first may be lost as they have time before the start of the recording to get accustomed to the interior.

To capture what a person looks at first, a suggestion is to perform the calibration outside the car and letting the participants start the evaluation by opening the door. Then the persons don't have any chance to get accustomed to the interior before the recording has started. Another advantage in starting the evaluation outside the car is that there will be results on where the person looks when entering the car. One result to obtain is if there is any interest in the Daylight Opening areas (DLO) when entering the car or if the persons looks at something special in the interior only visible from a side perspective (e.g. behind the steering wheel or under the seat).

If the calibration is to be performed outside the car it is still suggested to use the same distance of calibration as when evaluating the interior of the car and to keep a narrow placement of the calibration points.

Observations of the eye tracking data gave more information on which AOI's to use for the Main Study and where to place the markers in order to draw those AOI's.

When it comes to AOI's to use in the Main Study, a final suggestion has been made with a total of 12 AOI's (see Fig.34). The final decision on which AOI´s to include was decided in collaboration with experts within the department of Perceived quality. The biggest addition from the earlier used AOI's is to include an AOI in the backseat of the car too get an idea if the backseat is of interest in general or just for the five participants in the Pilot Study.

The placement, size and total number of markers is still a problem to consider but the effects from the markers have been greatly reduced and a suggestion has been made on how the markers should be placed (see Fig.34).

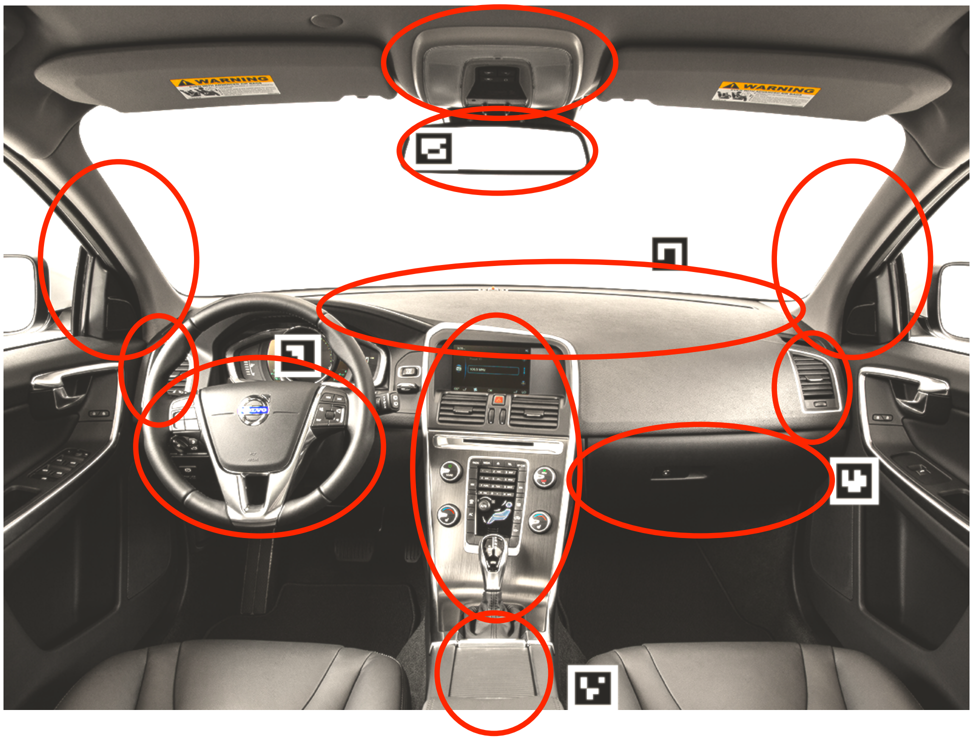


Figure 34: The final suggestion on which AOI's to use and how to place the markers in the Main Study.

The total number of markers has been greatly reduced in the final suggestion and the one with the highest recognition at the individual marker evaluation can be used.

Using observation in the study was proven to be a very valuable method which provided a lot of information. It is suggested to perform an observation even in the Main Study as one of the test leaders marks the glances during the evaluation on a picture of a car interior.

5.3.4 Exterior evaluation

Due to the many problems connected to an exterior evaluation it was decided in discussions with Volvo Cars that the evaluation of the exterior should be excluded. The main reason is in the eye-tracking systems limited performance and precision, leading to decreased validity and reliability.

One problem arising when observing the data is that there is a need to include a lot of markers to cover all the AOI's on all the sides of the car in the analyze of the exterior evaluation. The angles of the markers becomes very sharp and the recognition of the markers in exterior analyze is worse than for the interior analyze.

The poor recognition of the markers affects the possibility to draw AOI's in analyzes and it reduces the validity and reliability of the study. In order to get enough validity for exterior evaluation it is needed to evaluate the car in a static position. This means that the behavior of the participants needs to be controlled, which in turn makes the evaluation unrealistic.

5.3.5 Time plan

The time plan showed on an estimated mean total time for each participant to 30,5 minutes. This is an estimation only including the three main tasks (*Calibration*, *Recording of Eye tracking Data* and *Interviewing),* and there are also other tasks that need to be performed between the main tasks. It is also suggested to add some extra time as a buffer for unexpected events.

One thing that may reduce the total time is practice in the procedure of Calibrating. Practicing more will reduce the time needed. There is also a risk that there will be an increase of time for some persons if someone evaluates for a longer period of time than the mean Total time for Evaluation in the result.

It is suggested to focus on reducing the time as it is hard to motivate people to spend more than half an hour participating in a study.

5.3.6 The sample

The four different methods of constructing a sample showed on different advantages and disadvantages.

One important factor when considering a purchased sample is the eye tracking systems reliability. Eye tracking is a relatively immature technic which still suffers from problem areas. The pilot study and the manufacturer itself claims that one out of ten people are not anatomical suitable for performing an eye tracking study. The fact that it is impossible to determine before the actual test-person is in place and the calibration process begins, generates a risk when making an investment of a sample.

Using students at Chalmers was considered having a high impact on the representativeness of the sample since most of the students are in the age of among 20-25, which represents an almost un-existing segment of Volvo Cars purchasers. Furthermore, attracting students would probably requisite some sort of reward. One possible solution for reducing the investment was to take advantage of the curiosity of Volvo Cars newly launched product XC90 in order to attract people to attend the study without compensation. However, one potential problem of using such mediating tool is that the evaluation tends to shift towards areas and parts of pure curiosity instead of evaluating quality.

One possible way is to mix up the student sample with randomly picked elder people. This would however give rise to another problem, a need of using two different places for study to take place, which affects the validity of the study since the environment is changed.

The department of Ergonomics at Volvo Cars currently holds a pool of approximately 600 employees of volunteers; these employees are contacted for various types of studies but mainly for ergonomic research. The pool contains information about several important factors when creating a sample, such as age, sex, department of work, use of eyeglasses etc. Even though this sample only contains people currently working for Volvo Cars it can be hand-picked to reduce the risk of people being bias. In general, employees actively working with development of parts will be excluded, whereas employees from departments such as human relations and finance are more likely to be included.

Due to the limitations in time and budget it was decided to use the employees at Volvo Cars included in the pool of volunteers provided by the department of Ergonomics.

5.3.7 Build up

Other discussions while working in the process have been held on where to perform the study, what type of car to use as a mediating tool, how to improve the markers, and how to present the end result.

The venue to use

The general light in the analyze-lab is very good but there is a problem with direct light causing reflections as there is no cover over the fluorescent light sources. The analyze-lab offers a lot of space giving the chance for good movement around the car but a negative factor is that people tend to walk through the area which can act as a distraction element during evaluation.

In only focusing on interior evaluation, good movement around the car isn't necessary, but it is still suggested to find a venue more secluded and with covered light sources providing good general lighting.

The car to use

During the project it was discussed which car to use in analyze. A new car (The new XC90) would motivate people to participate in the study and act as a reward. At the other end it may mean that people would look at all new things and not focusing on the actual quality of the car, and it is therefore suggested going with an older car.

Another issue is that the project does include the evaluation from a premium perspective and it should therefore be a model of the upper end in the Volvo Cars portfolio with a high end specification package.

Making the markers more matte

To increase the markers recognition, by making them more matte, it was chosen to go with the method of cutting out matte black paper and to glue that on the markers. This is a time consuming method but it showed great improvement and offered stability to the markers. The other two methods, spraying and laminating, didn't provide enough improvement of the surface.

5.3.8 The visual presentation of the result

The work with the eye tracking system in the Pre- and Pilot Studies gave ideas on how the result from the eye tracking data can be presented visually.

It is important that the visual presentation of the result can be easily interpreted and mediated. The presentation should be possible to bring around between meetings to assist in discussions where PQ is to highlight areas of importance to other departments.

The final decision is to adapt the method of Heat Map used in D-lab 2.5 and to include AOI's graded according to attention with a color scale. The Heat Map can't be constructed in D-lab 2.5 as it would mean that one Heat Map for each participant is needed to be created. It is therefore decided to calculate the ARATIO of all participants and AOI's and manually construct the Heat Map with a picture of the interior and graphic software (e.g. Photoshop).

5.4 Answer to RQ2 and Conclusions

The Pilot Study was performed in order to answer to RQ2.

RQ 2: *How is a customer oriented study of Perceived Quality of cars using eye tracking technology constructed?*

SRQ 2.1: *What are the problem areas affecting the validity and reliability of a conducted study?*

SRQ 2.2: *Which qualitative data collection method should be used to increase the understanding of the eye tracking data?*

The following condensed bullet point list presents the conclusions made during the Pilot Study. These conclusions answering the research questions stated, the different topics of discussion are serving as headlines for the conclusion (*Quantitative, Qualitative, Interior Evaluation, Exterior Evaluation, Time Plan, Build Up, Sample,* and *Visual Presentation of the Result*).

Quantitative

- It was decided to use only two glance metrics in answering the questions of most and first as the other was seen as superfluous.
- Attention Ratio, ARATIO, will be used to determine what a person looks at the most and Time to First Glance, TFG, will be used to determine what a person looks at first.

Qualitative

- It was decided to construct a structured and well planned "Study Guide" to provide guidance throughout the process of performing the Main Study.
- It was decided to investigate the possibility of categorizing the participants according to the product aspects they mainly focuses on when talking about quality of the car.
- The qualitative interview needs to be better constructed and it is decided to consult experts within the area to obtain a better structure of the interview.

Interior Evaluation

- It is decided to perform the calibration outside the car and to record the first look into the car when the participant opens the driver's door.
- It was decided to use 12 AOI's to analyze the data from the interior evaluation.
- It was decide to use 5 markers to analyze the data from the interior evaluation. It was chosen to use the five best performing markers in the Marker Evaluation.

Exterior Evaluation

- It was decided to exclude the exterior evaluation from the Main Study due to the problems arising with marker recognition and movability reducing the validity of the result.

Time Plan

- The estimated time of 30,5 minutes per participant were decided to be too long and needed to be reduced as it is difficult to motivate for people to participate more than half an hour.
- With excluding the exterior evaluation the time will be greatly reduced but it should still not exceed 30 minutes of total time for each participant.

Sample

- It was decided to use a sample constructed by people in the pool of volunteers provided by the department of Ergonomics at Volvo Cars.
- This decision was made due to the low budget and tight time limit.

Build Up

- It was decide to find a new, more secluded venue for the Main Study. The new venue should provide good general lighting but no direct light sources visible that could disturb the recognition of markers.
- It was decided to not use the new XC90 as a mediating tool as it would be to much new features to investigate instead of the actual quality. But it is still needed to be a car that is perceived as Premium.
- In order to make the markers more matte it was decided to cut out matte black paper and to glue own markers. The time it would take is greatly rewarded in reduced risk of reflections in analyze.

Visual Presentation of the Result

- It was decided to use the method of Heat Map to visually present the result of the quantitative data from the eye tracking.
- The Heat Map is decided to be manually created with the values of each AOI's ARATIO obtained in D-lab 2.5.

6. MAIN STUDY

This chapter of the report will describe the method of work, display the results and discuss the results of the Main Study. The focus is on whether a study using eye tracking can be used to draw conclusions on where people look the most and first, together with arguments on why, while evaluating the quality of a car. This will be done by trying to find answers to the third research question, and subsequent questions.

RQ 3: *What does a person look at most and first when evaluating the quality of a car?*

SRQ 3.1: *What is the reason to that people look where they look?*

SRQ 3.2: *Are people aware of where they look?*

6.1 The methodology of performing the Main Study

The Main Study was structured and based on the knowledge gained through executing the Pre- and Pilot studies. The methodological procedure for each participant is reported as well as the methodology of how the qualitative interview was structured and how the questions were formed.

6.1.1 The build-up

The main study took place during three weeks in a clinic venue at Volvo Cars named TP5. TP5 is a venue normally used for testing and evaluation of cars, and stands up to the demands in light and seclusion discussed in the Pilot Study.

From the pool of volunteers provided by the department of Ergonomics 154 people were asked to sign up for a 30 minute study, and 25 persons responded and appeared for the study. The final sample consisted of 13 men and 12 women in different ages.

The sample of the study was chosen to simulate potential customers to Volvo Cars. Since people in different ages have different probability to purchase a car, the chance of being chosen in the study was adjusted accordingly. The target sample for this study is based upon market research performed by Volvo Cars in cooperation with NCBS (see table 6) and the distribution of ages of the asked participants can be seen in table 7.

Table 6: **The distribution of ages in NCBS**

Born	Percent
1980-1987	9
1970-1979	17
1960-1969	28
1949-1959	27
65+	19

Table 7: **The distribution of the asked participants**

Born	Amount	Percent
1980-1987	12	8
1970-1979	46	30
1960-1969	69	45
1949-1959	27	17

The car used in the main study was a Volvo XC60 with the highest specification available to assure high premium standard (see Fig.35).

Figure 35: The Volvo XC60 used as the mediating tool in the study

To increase the validity and reliability of the study, the test leaders aimed to perform each individual study under as similar conditions as possible. To ensure this, each of the test leaders performed the same tasks throughout the entire study, with one performing all the questioning and interviewing, and one performing all handling of the eye tracking system, including calibration and recording.

The entire procedure throughout the study was written down in the Study Guide (see appendix 5), this to provide structure and support to the test leaders with little experience. It was important that everything was planned carefully and executed as similar as possible to offer high validity to the results, and to minimize the risk of any unexpected events that will slow down the work.

The calibration of the system was performed with the help of a “calibration station” consisted of a chair offering good neck stability and five points attached on a white background (see Fig.36). The five points were made in black and white colors to give good contrast and to ease for the persons to focus on them while calibrating.

Figure 36: The calibration station used in the Main Study

6.1.2 Performing the Main Study

When the participant arrived at the clinic venue he/she was introduced to the study, what an eye tracking system is, and what was expected of him/her during the procedure of the study.

After the introduction some questions were asked regarding the person's relation to cars and Volvo, and then he/she were asked to sit down at the "calibration station" and to focus on the center point on the wall. The calibration was performed according to the calibration wizard (Ergoneers, 2011) and when done the participant was asked to stand up facing the driver's door.

Before the recording of the eye tracking data could be started, some key instructions were repeated and clarified. When the participant showed on understanding, the recording was started and the evaluation could start by opening the door and entering the car.

The free evaluation of the car was performed until the participant communicated that he/she was done and then the recording was stopped as there were no time restrictions.

During the free evaluation notations were made on where the glances were focusing by marking areas with a pencil on a picture of an interior. An example with markings can be seen in figure 37.

Figure 37: The interior picture used to mark attention manually while recording eye tracking data

When the free evaluation of the car was performed the participant was asked to stay seated in the car while the interviewing test leader entered the passenger seat to perform the interview. The test leader was equipped with the interview manuscript, a mobile phone with sound recording and the picture of the interior with marks of glances.

When the interview was done more information regarding the study was told, its goal and which attribute ordered the study.

6.1.3 Creating the qualitative interview

In the creation of the questions for the qualitative part of the study, there has been cooperation both with scientific experts from Chalmers University of technology as well as the department of Perceived Quality at Volvo Cars. It was important to rank the importance of the stakeholders and to keep in mind that the project was led by Volvo Cars and the department of Perceived Quality, and their opinions needed to be highest priority backed up with knowledge from the scientific expert.

The following description of how the interview and questions was structured and formed led to the finalization of the Interview manuscript inserted into the Study Guide, which can be found in appendix 5.

It was decided to divide the interview into four different parts of different level of detail and structure.

First part focused on adding some information not provided by the volunteer database. The questions concerned which department at Volvo Cars the person currently worked at, for how many years they had been at Volvo, their current car and earlier owned cars. Due to the simplicity of the answers needed, these initial questions also served as a “warm-up” in order for the participants to get comfortable with the situation.

The second part was the start of the actual interview regarding the quality of the car and here it was decided to start with open ended and wide questions, this to allow for elaboration as much as possible before interfering with more in depth and detailed questions. The first question, which was asked to get closure on the free evaluation, was asked to see the grading of the total interior quality of the car on a scale from 1-10.

During the rest of the second part of the interview the test-person were invited to reveal his or hers personal thoughts of prioritized areas of the interior. This part of the interview included three questions also aimed to elicit the participants’ opinion on which area of the interior that they evaluate *most frequently* and *firstly,* but also which areas that they perceive as *most important*.

Question 1: “*What are you looking at most when you are to evaluate the interior of a car?*”

If a certain are wasn’t expressed to be looked at the most, a subsequent question was put. In the occasions where a participant couldn’t name or define the area which was of interest, the test leaders guided in setting a name on the area the person indicated.

Subsequent question 1.1: *“Is there any certain area in the car which you are looking at more than other areas?”*

Question 2: “*What are you looking at first when evaluating the interior quality of a car?”*

Question 3: *“What areas do you think is of greatest importance when evaluating the interior quality of a car?”*

These three open ended questions was asked to provide answers that could be directly correlated with the results calculated from D-lab on what the sample looks at *most* and *first*, and to discuss the *importance* of those areas.

During the third part of the interview, the questions were structured to go more into detail and to encourage focusing on one area of interest at a time. This part was semi-structured and there were big differences between how discussions were engaged into freely regarding perceived quality topics or if there were need for additional guidance into that topic from the interviewer.

The fourth, and last, part of the interview was structured to summarize the interview. The questions was to see if the car met their expectations, if they would like to change their grading on the scale of 1-10, and what they thought people in general sees as important areas when evaluating the interior quality of a car.

6.1.4 Analyzing the data collected from the eye tracking

The data collected in the main study was analyzed in D-Lab with the use of Areas of Interest and calculations of glance metrics. The collected data was structured with all of the recordings, one for each participant person, sorted under one collective "project" (see Fig.38).

- Project – Main Study Interior
  - Subject 1: TP 1
    - Recording
  - Subject 2: TP 2
    - Recording
  - ....
  - ....
  - Subject 25: TP 25
    - Recording

Figure 38: Picture how the expanded tree in D-lab 2.5 is structured

First step was to draw the decided AOI's to be used in the Main Study (see Fig.39). The AOI's were all drawn in the Reference recording and made to be project AOI's, meaning that the AOI will be visible in all of the subjects where the marker controlling the AOI is visible.

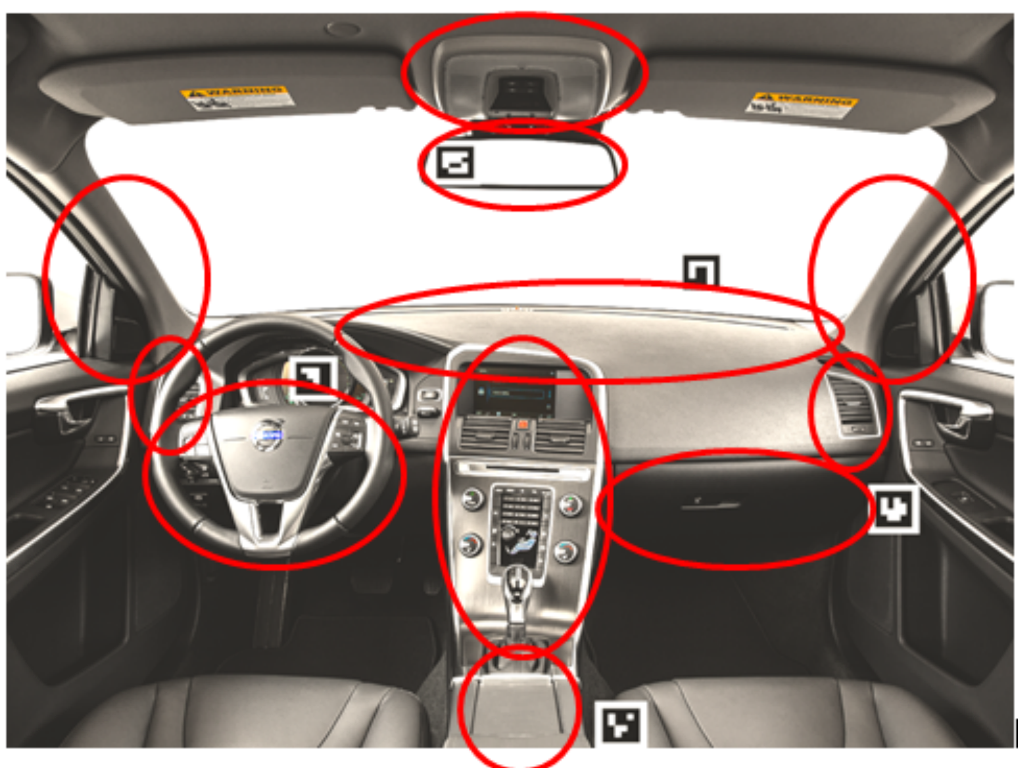


Figure 39: The 12 Areas of Interest (AOI's) used in the Main Study

Two different studies were created in D-lab. The first study, named “Main Study Calculations”, included the 12 different AOI’s earlier displayed in the report, while the second one, named “Heat Map Study”, included 23 different AOI’s (see appendix 6).

Both studies were prepared for analyze by calculating glances and then eliminate blinks and cross troughs using the default settings of 120 ms.

Calculating glance metrics and validity indexes

Glance metrics were calculated from both studies. Glance metrics calculated in the “Main Study Calculations” were Time to First Glance (TFG) and the Attention Ratio (ARATIO). These values were documented to offer data in order to answer the research questions.

The glance metric of Time to first glance (TFG) were used to display which AOI that D-lab registered to be glanced at first. By looking at the AOI for each participant with the lowest number in TFG, a list was summarized to display which AOI that was looked at first.

The glance metric of Attention Ratio (ARATIO) were used to display which AOI that D-lab registered to be glanced at the most. By looking at the AOI for each participant with the highest mean ARATIO, a list was summarized to display which AOI that was looked at the most.

From the “Heat Map Study” the Attention Ratio (ARATIO) was calculated to give data for construction of a heat map over the interior. It was decided to not use the built in Heat Map tool in D-lab 2.5 as it was only possible to create one Heat Map for each person and it wasn’t possible to create one showing the mean result of all 25 persons.

In addition to the glance metrics calculated, validity indexes provided from D-lab were documented to offer data to validate the result from the eye tracking collection. The values documented were the *full video marker validity* and the *full video eye validity* of all recordings.

- *Full video marker validity* – This index calculates the percentage of time during the recording where at least one marker is detected by the system.
- *Full video eye validity* – This index calculates the percentage of time where the eye-camera is successfully detecting the pupil and the corneal reflection.

The marker- and eye validity indexes provide information about how precise the actual recording is. A low index indicates that the precision of the eye tracking data can be questioned.

Calculating Total Score of Attention

In order to establish which of the AOI’s which were most viewed during the free analyze it was decided to use two different values. First value used was the ARATIO obtained from D-lab, which showed on AOI looked at the most. The different ARATIO for all test persons was summarized and then divided with the total number of test persons (n) to give a mean value of the ARATIO for each AOI (see formula 3).

$$Mean\ ARATIO = \frac{\sum ARATIO}{n}$$

Formula 3: The formula used to calculate the Mean ARATIO

A complementing calculating method was used, Total Score of Attention (TSA), in order to easier find metrics in the data clearly differentiating from the pattern.

First step was to establish a differential score for each AOI and participant. In taken the ARATIO for each AOI and take it to the power of two, a value named Score of Attention (SA) was calculated (see formula 4).

$$SA = ARATIO\hat{}2$$

Formula 4: The formula used to calculate the Score of Attention

All values of SA was then summarized and divided with the total number of participants in the sample (n) to give a Mean Score of Attention (TSA) for each AOI (see formula 5).

$$Mean\ TSA = \frac{\sum SA}{n}$$

Formula 5: The formula used to calculate the Mean Total Score of Attention

Observation of the eye tracking recordings

To validate if the calculated TFG from D-lab could be assumed to be the area which was looked at first, a manual observation of the recordings were performed as well.

The two test leaders did manually look through the recordings and decided which area each person looked at first.

6.1.5 Analyzing the data collected from the interviews

The sound recordings from the 25 interviews was listened through, transcribed word for word, and printed out on A4 paper. Each transcribed interview was gone through by each test leader as well as together to find relevant comments within the topic. When a comment of relevance was found, it was written down on a post-it note with some complementing information (see Fig.40).

The complementing information consisted of the test person which stated it, which area (if any) it concerned, the categorization of the comment (e.g. Function, Material, Layout/Design etc.), and a short description of the question leading to the comment and at which level of probing the comment was expressed.

The level of probing, marked with an encircled number on each post-it, was used to easier estimate the importance of that comment.

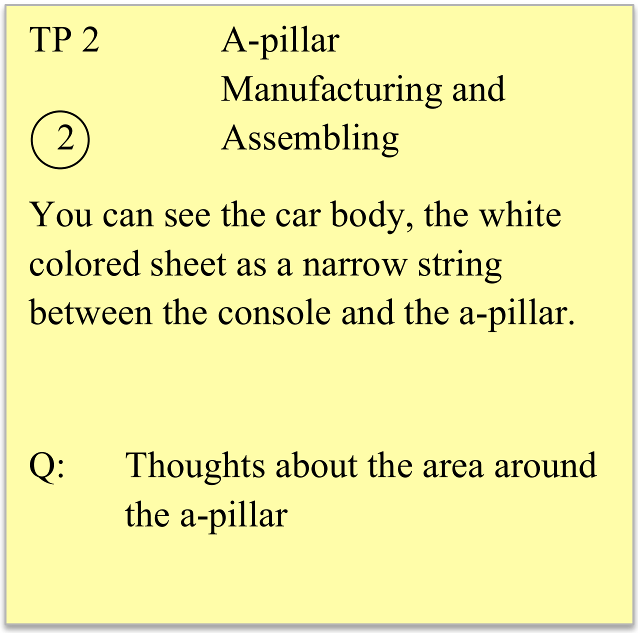


Figure 40: An example of how the post-it notes were constructed

The coding of level of guidance was decided to use in order to differentiate the weight of similar answers as they could were more or less self-thought of. The first level was when the comment was expressed without any guidance and just in discussion of the overall quality of the car. Level two was when the participant was guided to talk about a certain AOI in the car. And the third level was when probing was needed to elicit a topic of interest, as when guided to talk about specific Perceived Quality topics (e.g. Split lines, material etc.).

The post-it notes created were grouped and organized by attaching them on a wall connected to different topics which the post-it notes concerned. The organization and structuring of the comments was used in the work to see correlations, and to compare with the eye tracking data, as well as given support in answering of the research questions.

6.2 The result from the Main Study

This chapter will display the results from the Main Study on where people look the most, and first when evaluating the interior quality of a car.

6.2.1 Result on where people glanced the most

The result on which area the test people glance at the most was decided from the Attention Ratio obtained within each AOI together with the participants own perception and what they expressed to be the answer in the interviews answering question 1.

Question 1: "*What are you looking at most when you are to evaluate the interior of a car?*"

The eye tracking data on where people glanced the most

The result of the mean Attention Ratio shows that the *Center Stack* were most viewed with 14,186 percent followed by the *Steering Wheel (7,462 %)*, and that the *Center Console* were least viewed with 0,430 percent. All of the AOI's with marked placement in attention can be seen in figure 41.

The Mean Score of Attention didn't follow the exact same order as the ranking in mean attention ratio (see table 8). The big differences was *Right Air Vent* which placed 7th on mean attention ratio but only 10th on mean score of attention, where both *Rear View Mirror*, *Left A-pillar*, and *Back* got a higher score than *Right Air Vent*. The result shows on a differentiation in the pattern between participants when looking at some AOI's. The result in Mean Score Attention of *Back* and *Left A-pillar* indicates that one or few persons looking unusually much on those areas.

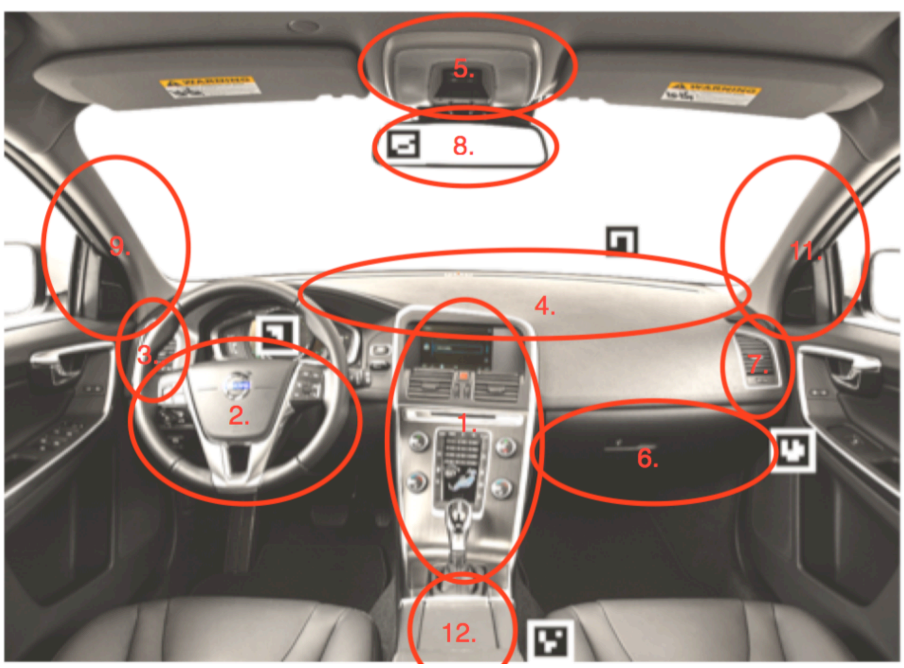


Figure 41: The 12 AOI's with their ranking position in attention

Table 8: **The results in attention of the 12 AOI's**

Rank	Area of Interest (AOI)	Mean Attention Ratio (ARATIO)	Mean Score of Attention (SA)
1	Center stack	14,186	247,904
2	Steering Wheel	7,462	78,328
3	Left Air vent	3,048	14,434
4	Top Instrument Panel	2,338	9,573
5	Ceiling	2,078	8,185
6	Glove Compartment	1,428	2,906
7	Right Air Vent	1,034	1,694
8	Rear View Mirror	0,953	1,841
9	Left A-pillar	0,839	2,169
10	Back	0,564	1,987
11	Right A-pillar	0,576	0,728
12	Center Console	0,430	0,581

The interview data on where people glanced the most

In order to gain a summary of the qualitative data from the interview, each answer on the interview question was gone through to see how many who mentioned each AOI (see table 9). Observe that a participant could mention more than one AOI when answering the question.

Table 9: **The results of what areas the interviews shows to be looked at the most**

Rank	Area of Interest (AOI)	Inclusive in the pre decided 12 AOI's	Number of participants mentioning the AOI as an area which they looked at the most
1	Steering Wheel	YES	10
2	Seat	NO	10
3	Center Stack	YES	9*
4	Door	NO	4
5	Top Instrument Panel	YES	3
7	Driver Environment	NO	2
8	DIM	NO	1
9	Gearshift	NO	1

* When someone expressed that they looked at the "stereo" it was marked as "center stack".

Another result from the interview is all the answers not referring to a specific area within the interior but instead a product aspect which was most looked at when evaluating the interior quality of the car. The product aspects mentioned are summarized in table 10. Observe that a participant could mention more than one product aspect when answering the question.

Table **10:** The results of which product aspects that was mentioned to be looked at the most

Which product aspect expressed to be evaluated at the most during the evaluation	Number of participants mentioning the product aspect as an area which they looked at the most
Fit and Finish	8
Feel in Material	6
Controls and Buttons	5
Matching Material	4
Simplicity	4
Overlook	3
Solidity	3
Stitches	2
Layout/Design	2
Seating Comfort	1

Correlation between eye tracking data and interview data

Looking at the data collected in the main study, there were differences in what the interviews expressed to be looked at the *most* compared with the actual data collected with the eye tracking system.

In the results from answering the interview question there is three AOI's more mentioned in the answers than other AOI's, and these are the *Steering Wheel*, *Center Stack*, and the *Seat.* The Center Stack and the Steering Wheel is both placed at the top in the list of AOI's when calculated by D-lab to have the highest Mean Attention Ratio.

The *Seat* isn't included in the pre-decided AOI's, meaning that the eye tracking data from D-lab didn't include that area. Other areas mentioned during the interview session and not included in the pre decided AOI's is the *Door* which were mentioned by 4 persons, *Gearshift* (1 person), and *Dimmer (1 person)*.

Two persons couldn't decide on a specific area where they looked the most and answered the *Driver Environment*.

When looking at participant by participant the result is very different (see appendix 7). Only 6 out of 25 persons (24%) saying that they are looking mostly at the area which the eye tracking data shows. The summarized result at the other hand indicated some correlation as, except for the *Seat* which were mentioned 10 times but not analyzed in D-lab, the two highest scoring AOI's from the eye tracking data, *Center Stack* and *Steering Wheel,* were positioned at top even in the result from the interviews.

One result shown was that there's a mix-up between the two AOI's *Steering Wheel* and *Center Stack*, meaning that when the calculated Attention Ratio showed on *Center Stack,* the interview showed on *Steering Wheel*, and the other way around. The amount of participants where this mix-up was shown was 7 out of 25 (28%).

Visual presentation of the result on where people glanced the most

One of the most important deliveries in the project is an easy visualized presentation of the result on where people look, to give the employees at Perceived Quality at Volvo a useful tool in their work.

It was decided to use a heat map drawn from the Attention Ratio from each AOI obtained from D-lab. In the heat map more Areas of Interest was included and the total number of AOI's ended up to 23. A clear description of each AOI and the mean ARATIO each achieved can be found in table 11.

The color scale used in drawing the heat map is divided in 30 gradient steps (see Fig.42). The highest ARATIO is 3,31616 and that is for the Center Stack. This means that the interval of the scale ends up to be between 0 and 3,31616. The 30 steps of the color-scale is displayed in figure 42.

Step 1 ---Step30
0 -- MEAN ARATIO ------------------------------------3,31616

Figure 42: The color scale and the step size used to make the Heat Map

To establish which color-code each AOI should be colored with, the highest value of 3,31616 is divided by 30. This means that each step is 0,1105 %. Each AOI's ARATIO is then divided by the step size of 0,1105% resulting in a step value. This step value is to be rounded to the closest integer and gives which step each AOI belongs to and the connecting color-code (see table 11).

Table 11: **The result of the attention of the 23 AOI's building the Heat Map**

Area of Interest (AOI)	Abbreviation	Mean Attention Ratio (ARATIO)	Step	Color Code
Center Stack	CS	3,31616	30	#FF0000
Center Air Vent	CAV	3,074	28	#FF2300
Stereo Buttons	SB	2,72232	25	#FF5700
Center Steering Wheel	CSW	2,7394	25	#FF5700
Glove Compartment	GC	2,43964	22	#FF8C00
Center Display	CD	2,213	20	#FFAF00
Top Instrument Panel	TIP	1,64124	15	#F7FF00
Ceiling	C	1,48132	13	#D4FF00
Left Steering Wheel Buttons	LSWB	1,17256	11	#B0FF00
Rear View Mirror	RVM	0,8546	8	#7CFF00
Right Door	RD	0,8472	8	#7CFF00
Left Air Vent	LAV	0,86132	8	#7CFF00
Right Steering Wheel Buttons	RSWB	0,70464	6	#58FF00
Left Door Handle	LDH	0,70172	6	#58FF00
Gear Shift	GS	0,70416	6	#58FF00
Back Center Console	BCC	0,37076	3	#24FF00
Left Bliss Housing	LB	0,3378	3	#24FF00
Left A-pillar	LAP	0,34204	3	#24FF00
Right A-pillar	RAP	0,24028	2	#12FF00
Right Bliss Housing	RB	0,20236	2	#12FF00
Front Center Console	FCC	0,16012	1	#00FF00
Left Door Buttons	LDB	0,16032	1	#00FF00
Right Air Vent	RAV	0,16344	1	#00FF00

Including the correct color-codes and coloring the interior picture resulted in the finalized Heat Map of the interior of the Volvo XC60 (see Fig.43).

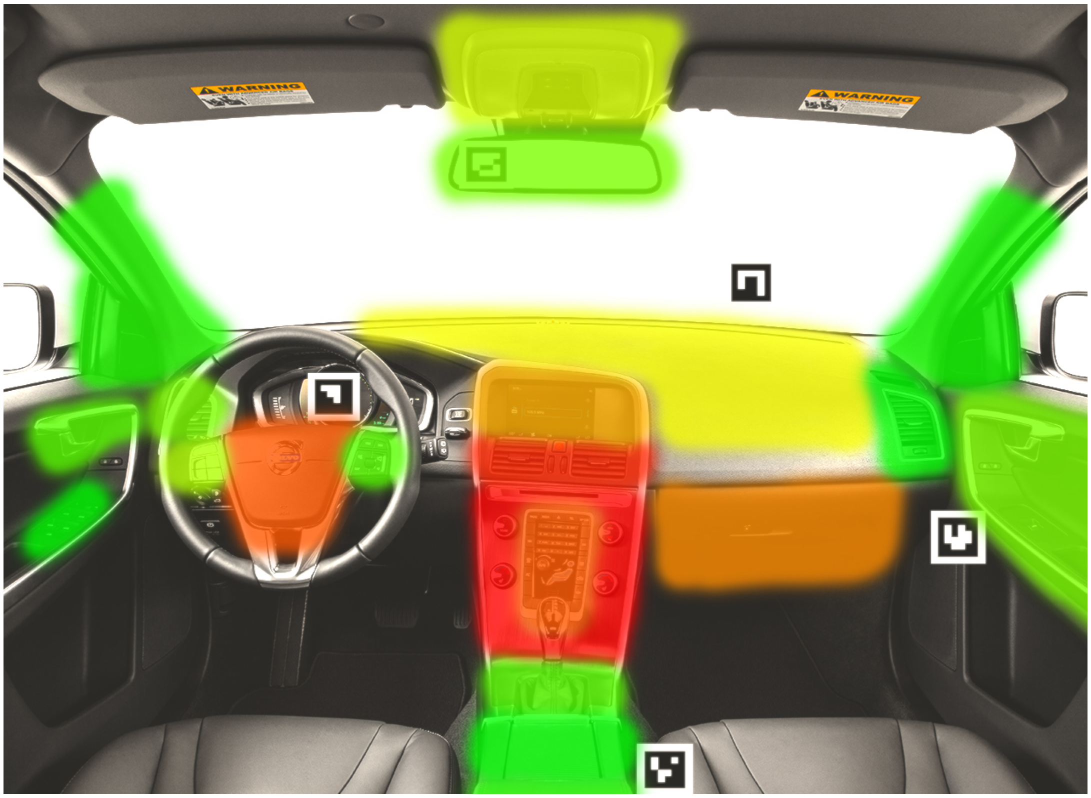


Figure 43: A miniature of the finalized heat map over the interior of the Volvo XC60 used in the study.

6.2.2 Result on where people glanced first

The result on which area the test people glance at first was displayed through three different methods. First method was the Time to First Glance value obtained for each AOI, the second method through manual observations of the eye tracking recordings, and third the participants own perception through what they expressed to be the answer in the interviews answering question 2.

Question 2: "*What are you looking at first when you are to evaluate the interior of a car?*"

The result from D-lab analyze tool, the manually observed results, and the comments from the interviews is displayed in table 12.

In evaluation, none of the participants paid any attention to the DLO area around the driver's door when opening it.

Table 12: The result of the areas looked at first

TP	First area viewed according to TFG	First area viewed according to manual observation	First area viewed according to comment from interview
1	Center stack	Seat	Seat
2	Steering Wheel	Steering Wheel	Seat
3	Top Instrument Panel	Center stack	Seat
4	Steering Wheel	Dimmer	Seat
5	Left Air Vent	Dimmer	Center Stack
6	Steering Wheel	Center stack	Seat
7	Center stack	Steering Wheel	Steering Wheel
8	Steering Wheel	Steering Wheel	Steering Wheel
9	Top Instrument Panel	Steering Wheel	Center Stack
10	Steering Wheel	Steering Wheel	Center Stack
11	Steering Wheel	Steering Wheel	Center Stack
12	Center stack	Center Stack	Center Stack
13	Center stack	Steering Wheel	Center Stack
14	Steering Wheel	Steering Wheel	Seat
15	Center Console	Door	Center Console
16	Left Air Vent	Center Stack	Center Stack
17	Back	Door	Center Stack
18	Steering Wheel	Steering Wheel	Overlook
19	Steering Wheel	Center Stack	Center Stack
20	Left Air Vent	Dimmer	Overlook
21	Center stack	Steering Wheel	Dimmer
22	Glove Compartment	Steering Wheel	Seat
23	Left A-pillar	Steering Wheel	Overlook
24	Steering Wheel	Door	Steering Wheel
25	Left Air Vent	Door	Overlook

Correlation between calculated TFG and eye tracking data

First step in analyzing the correlation is to compare the calculated TFG with the manual observation looking through the eye tracking recordings. All data regarding what was looked at first can be found in table 12.

In order to get a brief picture of how the distribution between the AOI's was in first glance, see table 13.

Table 13: **The summarized result of the areas looked at first from the three different methods**

AOI	TFG	Observation	Interview
Steering Wheel	10	12	3
Center Stack	5	5	9
Seat	0	1	7
Left Air Vent	4	0	0
Top Instrument Panel	2	0	0
Left A-pillar	1	0	0
Center Console	1	0	1
Glove Compartment	1	0	0
Back	1	0	0
Door	0	4	0
Dimmer	0	3	1
"Overlook"	0	0	4

The amount of participants which had an exact correlation between the AOI calculated with TFG and the manual observation were 7 out of 25 (28%).

The amount of participants that had a "near" correlation was 5 out of 25 (20%). "Near" correlation means that the two AOI's is adjacent placed in the analysis (e.g. the Left Air Vent and the Steering Wheel) (see Fig.44).

Example: If the result from TFG shows on the Steering Wheel but comments state the Left Air Vent, then that's called a "near" correlation as the areas are adjacent.

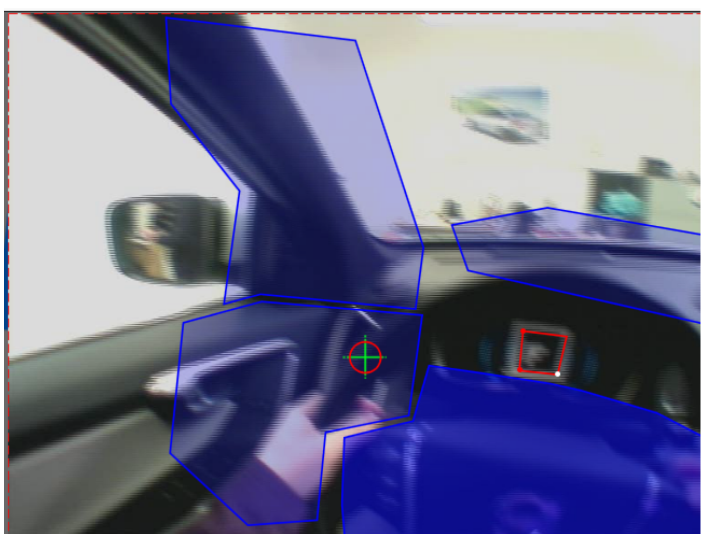


Figure 44: Showing AOI's closely placed which may result in "near" correlation (Print screen from D-lab 2.5).

Another result shown, as in the analyze of where people look the most, was that there's a mix-up between the two AOI's Steering Wheel and Center Stack, meaning that when TFG showed on *Center Stack,* the observations showed on *Steering Wheel*, and the other way around. The amount of participants where this mix-up was shown was 5 out of 25 (20%).

Some areas which were decided through observation to be the area which to be looked at first weren't included in the pre-decided AOI's. These areas were the *Seat* where one person looked at the first, the *Dimmer* (3 persons), and the *Door* (4 persons).

Correlation between interview comments and eye tracking data

Second step is to see if there is any correlation between what people say they look at first and what the eye tracking data shows.

Two areas were over represented in the comment from the interviews, and that was *Center Stack*, mentioned by 9 persons, and the *Seat* (7 persons). The *Steering Wheel*, which in the observation was seen as the first area to be looked at for 12 out of 25 persons, didn't get mentioned more than 3 times in the interview session. At the other hand was the *Seat* mentioned by 7 persons, but the manual observation showed that only 1 person looked at the seat first.

Some participants did have problem deciding on one certain area which they looked at first and 4 out of 25 (16%) did leave the answer that they made a sweeping overlook over the entire interior first thing when evaluating the quality.

6.2.3 Results on the comments from the interview

In this section of the result the comments concerning each area of interest are summarized to offer an overall view on what is seen as important in each and every of the predicted AOI. The chosen comments cited in the section is chosen to offer a descriptive comment to the text in each section.

The comments were sorted according to the area it concerned and due to the nature of some of the questions asked during the interview it was also decided to include two additional groups, "holistic view" and "previous experience" (see Fig.45). These groups were aimed to capture statements regarding the overall view upon quality not connected to a specific area of the interior.

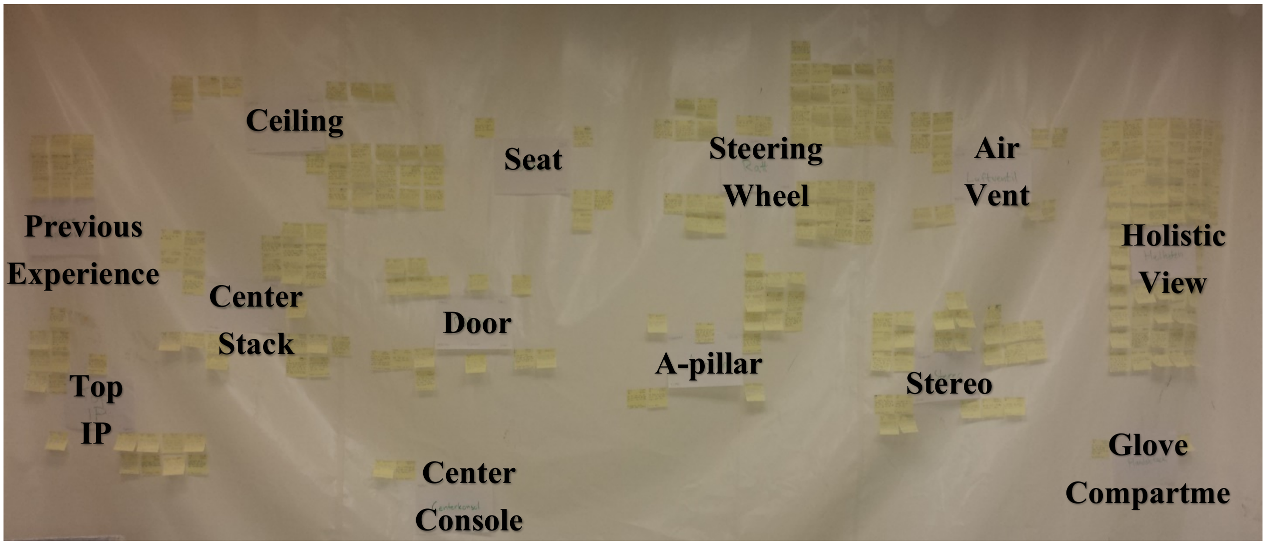


Figure 45: Picture of the comments grouped in 11 groups on a large white wall

The different topic areas will be discussed in order from how many comments of relevance each topic had in the sortation with the two more general topics at the end.

1. Center Stack (Center Stack, Stereo and Air Vent)
2. Steering Wheel
3. Ceiling (Ceiling and Rear View Mirror)
4. Top Instrument Panel
5. A-pillar
6. Other Areas (Glove Compartment, Door, Seat and Center Console)
7. Holistic View
8. Previous Experience

Center Stack

The Center Stack is the most viewed AOI and the topic which had most comments. The AOI called the Center Stack was including both the Stereo and the Central Air Vents and all comments concerning these two sub-areas were included in the results of the Center Stack.

The comments mainly concerned the complexity of the area and the chosen materials, more than the actual aesthetic layout and design. The complexity is primarily expressed through the amount of controls and buttons and that it needs to be easily understandable.

The material concerns include both visual and haptic comments. The visual concern is the differences in material selection between adjacent areas and also that the material perceived to differ between components thought of having the same material. The haptic concern was mainly that the feel of the material was poor and gave a cheap impression.

Some persons, even though a small percentage, talked about the fit and finish of components within the area of the Center Stack.

Steering Wheel

The second most viewed area was the Steering Wheel and without any probing at all there were only three persons that had comments on the Steering Wheel. The most comments was expressed first when the interviewer asked the interviewee to focus solely on the area of the steering wheel and when asking about lines, shapes and details there were many comments concerning different parts of the steering wheel.

When asking about the area of the steering wheel the comments were mainly concerning feel of material, functionality, layout/design, but also fit and finish. The feel of material was concerning the grip and it was highlighted that it is crucial to feel good as that is the component you interact with the most in the car.

"You are sitting and holding on to it all the time… it is important that if feels good to grip".

- Test Person 25

In the comments concerning functionality and layout/design, the comments concerned the buttons and the gear shift paddles, the buttons were expressed by some to look old but very functional, the gear paddles to be redundant, not providing any additional function. Both the buttons and gear shift paddles were expressed to disturb the overall visual impression.

During discussion regarding the fit and finish within the area of the steering wheel, there is overall positive comments which was expressed first at the third level of probing. Some participants highlights the split-line between the airbag and the rest of the steering wheel, as it the gap between the materials is wider than other split lines. Comments where also stated regarding the two folded design of the steering wheel and it was emphasized that the execution of such bold design needs to perfect in order to not disturb.

"If you are going to do something like this, then you have to make it good. Otherwise it just disturbs. Then it is better to just make an original designed steering wheel".

- Test Person 13

Ceiling and Rear View Mirror

The comments regarding the Ceiling panel mainly concerned the layout/design and that it feels big in comparison with the actual function.

Two specific components that got a lot of comments expressing that they do not harmonize with the rest of the premium execution, the extra-featured button to control the panorama roof and the city safety housing behind the rear view mirror.

There were also comments regarding the rear view mirror, many of the participants expressed that they had seen the new designed rear view mirror and made a direct comparison with that, stating that the new design is a big improvement in terms of design.

Top Instrument Panel

The comments regarding the Top Instrument Panel was mainly concerning the material selection, specifically the surface structure, and that the layout/design of the top instrument panel should be clean. The comments concerning material and surface are all positive, expressing that it gives a good impression, the negative comments are related to the functions on the top instrument panel, especially the Head-Up Display (HUD).

"I think the instrument panel feels good… it has such a surface that feels a little bit "more", it is probably plastic that as well but it feels more lavish, because it has more structure"

- Test person 9

"There is a lot of stuff upon the panel which should have been more hidden maybe, some of the functions…"

- Test Person 25

A-pillar

The comments regarding the A-pillars will be summarized together even though the left A-pillar did receive a higher ARATIO than the right one.

Comments on the area around the two A-pillars were mainly positive even though many participants stated that this area is complex with many materials and functions meet in relatively small area.

"It looks alright… good gaps between the instrument panel and the A-pillar."

- Test Person 24

Two participants identified a see through in the right A-pillar area, which they stated reduced the overall impression of the area.

The Bliss Housing is the one individual component, which by far gets the most negative comments in the study as 8 persons out of 25, gives comments such as, the plastic material do not meet the standard of the surrounding materials and components.

Other areas

There were also some, but few, relevant comments within other areas.

One area was the Glove Compartment, where two people commented on the importance of a good sound when opening and closing.

Another area was the Centre Console where three persons commented on the poor quality in the plastic material around the storage department.

The third additional area is the Seat which had a few comments on the comfort and appearance of the seat.

The fourth and last area is the Door which had more comments on material, mostly positive but some negative on differentiation in chromium.

6.2.4 Values concerning the validity of the study

The values regarding the validity of the eye tracking data were registered and used to analyze the validity of the study and the eye tracking system. Some values are greatly differentiating from others and it needs to be discussed on why that might be to give further tips to future research.

The different values (see table 14) affecting the validity and reliability of the test is Total time of recording (TT), Recognition rate (RR), Marker Validity (MV), and Eye Validity (EV). The full table of values can be found in Appendix 8.

The mean time used to evaluate the car is approximately 4 minutes. It should however be noticed that the evaluation time differs a lot between the individual participants, ranging from 2 till 22 minutes.

The Recognition rate also showed big discrepancies, ranging from 36-98 percent, with a mean value of 81. Big variances in recording quality were also noticed for the marker- and eye validity which differs from 43-92 and 46-95 respectively.

Table 14. Validity indexes from the Main Study

Mean Evaluation Time	Mean Recognition Rate (RR)	Mean Marker Validity (MV)	Mean Eye Validity (EV)
04:18	81,16	73,43	77,76

6.3 The discussion about the main Study

This chapter will first discuss *why* people look at certain areas and aspects by combining eye tracking data and statements provided during the interview. Furthermore is an attempt to categorization of the participants discussed. The chapter is concluded with a discussing regarding the validity and reliability of the study.

6.3.1 Why people look where they look

The results from the study has given answers of where people look the most and first when evaluating the interior quality of a car and in this section will discussions be held on why people look at certain areas more than other. In order to find answers analyze were made to find correlations between the eye tracking data and the interviews. Looking to the areas which were most viewed and the comments regarding those areas showed on similarities on what's draws attention.

When going through the comments it gets clear that some persons tend to talk more about some topics than others and that each person's answers is focused within certain areas. This led to discussions if potential customers can be categorized according to what they perceive to be of importance when evaluating the interior quality of a car.

The result showed clearly that the Perceived Quality of the interior is evaluated both with the eyes and the haptic senses. Some persons tend to use one method of evaluation more than the other but they are working together to give a total impression of the quality. The three categories discussed are all included in both the Visual-, and Feel Quality.

"It is both the visual and the feel when you are touching it."

- Test Person 4

When analyzing the results from the study it is suggested to use three main categories in categorization.

Functionality	For the person which looks to the functions in the car and that the quality is measured by the right functions available, that it is easy to use and understand, and that it feels good to use the controls and buttons controlling the functions.
Material	For the person which looks to the materials in the car and that quality is measured through the right materials, evaluated through both visual and haptic senses. The importance is put into that the selection of materials harmonizes together and that the chosen material feels premium.
Manufacturing and Assembling	For the person which looks to the manufacturing and assembling of parts and components and that quality is measured through the craftsmanship and solidity of the parts.

In the discussion the three different categories will be discussed from the perspective of the top placed areas in attention and what makes them most viewed. The terminology used in the interviews and stated in the comments will be connected to the attributes defined and divided in the Technical Perceived Quality (TPQ) created by (Stylidis et al., 2015) (see figure 46).

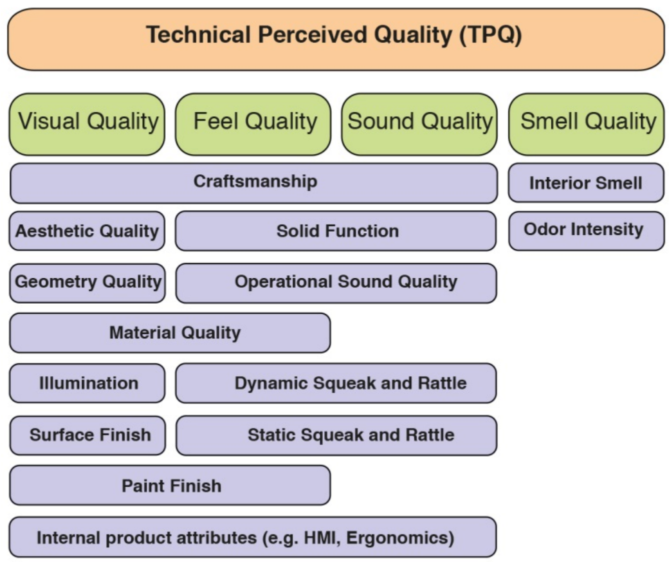


Figure 46: Technical Perceived Quality attributes (Stylidis et al., 2015)

Functionality

Functionality was a topic often commented when talking about the two most viewed areas, the Center stack and the Steering Wheel. Both areas include functions that are used daily with the clearest example in the steering wheel, but also functions controlling the temperature and the stereo located at the Center Stack. The parable between the two areas and that functionality is a topic which they have in common led to discussions if people look at an area because of the high possibility of interacting with the car and its different functions?

A reason to that functionality is often commented on may be that it includes the possibility to evaluate many attributes of the Technical Perceived Quality (TPQ).

The comments shows that the evaluation of the functions are both visually and through feel, and that people evaluate the fit and finish (*Craftsmanship*), the solidity in using the functions (*Solid Function*), that it makes the right sounds when using the functions (*Operational Sound Quality*), and that the material feels good when gripping it (*Material Quality*).

The more attributes possible to evaluate together with a high complexity due to many controls and buttons is assumed to demand longer time for evaluation. This is assumed in the whole interior of the car but especially noticeable when looking to the two top placed areas and their complexity.

The complexity of too many controls and buttons is assumed to affect the Visual Quality of both the Aesthetic-, and Geometry Quality, but also other attributes.

Functionality is also an area which offers comfort to the participants and makes them feel settled. This is showed when looking to the areas which are viewed first in evaluation with the seat and the steering wheel high placed, together with comments stating that the first thing to be done while entering the car is to adjust the chair and to feel the steering wheel. First when they are settled they can start to evaluate the quality of the car.

One topic mentioned in the interviews while talking about functions is that people tend to think that an important factor is that there should be no risk of feeling stupid when interacting with the car. Providing good Human Machine Interaction (HMI) is important and affects Visual-, Feel-, and Sound Quality included in the attribute *Internal Product Attributes*.

It is assumed that the result from the eye tracking is directly connected to functions of high complexity or that it is needed to visually evaluate what is happening while using a function. One example is the Air Vents within the area of the Center Stack, commented on by many to look old. The function of setting the direction of air gives clearly visible result and people tend to expect more from the function than just plastic parts moving when adjusting the angle.

Material

Material is a topic commented when talking about all areas within the interior, from the top placed Center Stack to last placed Center Console, and it can be connected to both the Visual-, and Feel Quality of the material. The many comments regarding material throughout the interviews led to discussion if people look at an area to evaluate the visual- and feel quality of the material within that area?

The chance of evaluating the Feel Quality of material is assumed to the eye tracking result. Evaluation of Feel Quality is assumed to be the reason to why the Steering Wheel has lower attention than the Center stack even though that the function of the steering wheel is the single most used function in the car. The evaluation of the steering wheel may be done through gripping it while looking at other areas. This is strengthened by comments stating that the feel of the steering wheel is important together with the low amount of negative comments on the visual appearance.

“Well it is the one thing you are sitting and holding on to the entire time… it is important that it feels good to grip”

- Test Person 25

The other way of evaluation, through vision, is mostly used when there are materials on components not often gripped or touched, e.g. the Top Instrument Panel. The comments are then mostly concerning that the materials selected needs to harmonize and match together and the harmonization of materials in the interior is considered to be one of the largest contributors to the holistic view of quality.

People tend to look at a certain material that stands out, for example a luxurious material (e.g. chromium or a structured material) and then evaluates to see if the other materials stand up to the performance of that material. If there is some materials not meeting the performance there are negative comments as the difference becomes obvious. Examples in the study are the plastic parts around the dimmer, central air vents, steering wheel buttons, and additional equipment (City Safe Housing and Bliss Housing).

It is important to understand that some material selections are specifications chosen by each customer, but there are standard materials which need to be assured to not differentiate too much in performance from the chosen materials.

The Visual-, and Feel Quality cooperates in creating an impression of the overall Material Quality and some components needs to be focused on given the right feel as some needs focus on give the right visual appearance, but it is safe to say that if some component or part stands out it will be noticed by the potential customers and give an negative impression. One of the participants did make a comment which clearly described the issue of different materials.

"It feels like the ones working with the Instrument panel is working for themselves, and the door panels for themselves, and that they really don't talk with each other. Here's a very clear difference in the color and structure of materials"

- Test Person 16

Manufacture and Assembling

A third topic of discussion when talking about different areas in the interior of the car is Manufacture and Assembling. The term is including fit and finish (*Craftsmanship*) and solidity (*Solid Function, Operational Sound Quality*, and *Squeak and Rattle*).

There are persons who clearly expressed early in the interview that fit and finish is an important area to evaluate, and some even lifts forward specific examples on bad fit and finish. The high result of persons saying that fit and finish is the thing they look at most together with comments on solidity led to discussions if people look at an area to evaluate the craftsmanship within that area, including fit and finish and solidity.

The high complexity resulting from controls and buttons within an area is demanding high performance in manufacturing and assembling to give and overall impression of quality. Even though fit and finish is shown to be the most important product aspect to look at there is not many negative comments. This is indicating that the overall execution of fit and finish in the XC60 is meeting the requirements of the customers.

One specific split-line which did get some negative comments was the one between the airbag and the steering wheel. This is a split-line which is much larger than other split-lines in the car and maybe that's a reason for the comments. It is shown from earlier discussion that a component which stands out from the rest will draw attention towards itself, and often contribute negatively to the overall impression. It is suggested that Craftsmanship is first noticed and commented on when faults are very obvious.

"If it would be a large gap here, or if you could fit your nail here but not here, then it probably would disturb me every day."

- Test Person 10

From the theory it is stated that Perceived Quality is assessed in a comparison context and almost the entire sample own, and are used to drive, a Volvo Car. It is assumed that the low amounts on comments of poor fit and finish is due to the low amount of things to compare with as their cars is probably at the same level or lower when it comes to quality aspects affecting their perception of quality.

If a person in other hand has seen a car which, according to PQ at Volvo Cars, is seen as very good in the perspective of perceived quality, he/she could use that in a comparison context and the result may have differed.

Fit and finish was seen to be hard to define and to evaluate, but it seemed easier to define and to comment on poor solidity. Evaluation was made both of Visual-, and Feel Quality, and it was seen that evaluation of solidity is strongly connected to the interaction with functions.

When looking at the type of comments on poor solidity in functions, some specific components where mentioned throughout the respondents. One was the air vents located in the center and one was the front pocket in the center console, both of them with negative comments on the Feel Quality of using the functions as it felt that they would break.

"If it feels like the buttons would brake within the next month, then you are getting worried."

- Test Person 21

The overall impression on why Fit and Finish, and solidity are important to evaluate is that people want the car to feel well thought thru. It needs to show that there have been effort put into making the car, manufacturing and assembling even the small details correctly, and this because of that the car is seen as a quiet expensive car with high expectations from the customers.

Other discussions on the result from the Main Study

The discussion on why a person's looks at a certain area more than other led to several additional hypotheses regarding other findings in the study.

Additional features may effect overall perceived quality negatively

Another discussion regarding functions was that additional features which can be selected were negatively perceived. The additional features such as the Bliss Housing, City Safety Housing and the control to the panorama roof all got negative comments with indications that it is possible that additional features are less prioritized in development. The negatively experienced design is assumed to draw attention and affect the holistic overall Perceived Quality negatively. There are comments with supposition that the additional features are added afterwards with less precision and integration than the original design.

There is also a problem with the materials in the additional features as it is often perceived as cheaper and that it doesn't harmonize with the surrounding materials. As it is now, the additional features affect the Perceived Quality negatively for some of the participants.

The higher the complexity the higher the attention

One common topic in the comments regarding all of the three categories is complexity. In discussion of functions complexity is clearest shown in the amount of controls and buttons, in material complexity is shown in many different materials which needs to harmonize and match together, and in the manufacturing and assembling it is shown in the many lines and shapes meeting in smaller areas (e.g. area around the a-pillar and air vent on both sides.)

The complexity within an area and the harmonizing between components are not only affected by the amount of components but also how easily perceived the area is. As previously described in the theory of eye tracking is total glance time within a certain area affected by how important that area is but also that the complexity is high, which demands more time for interpretation of that area.

One example is the button controlling the panorama roof, which is an area of low complexity. The button can quickly be interpreted meaning that low fixation time could both mean that the button is quickly decided to look bad or that a quick glance it's all it takes to decide that there is nothing disturbing in the area.

The interior quality in a car is perceived in a holistic perspective

One of the clearest conclusions which may be assumed is that people in general tends to look at the interior quality of a car from a holistic perspective. This is assumed to be due to the difficulty in defining and talking about quality, and especially Perceived Quality. Many of the participants often use words like nice when describing an area or component but when asked to describe why it looks nice they can't find the words.

The discussions that something which stands out draws attention and affects negatively give proof to the importance of a holistic view and there are comments stating that a single part or component can ruin the overall impression of quality.

"If there is something which bothers you when looking, then I think the overall impression is destroyed quite fast."

- Test Person 20

"The holistic impression is important… it can look very nice, but everything can be ruined fast.

- Test Person 13

6.3.2 External validity

It is difficult to draw any detailed conclusions on how well the chosen sample represents a population of potential customers of a car. However, from a scientific point of view is the fact that all of the participants currently work at Volvo Cars the main reason why the external validity should be considered as highly limited.

The major topic of concern according to the authors is that almost the entire sample is driving a Volvo as their primary vehicle today. As mentioned is perceived quality often judged in a comparison context and the lack of ownership of competitive products can raise a concern over the participants' ability to evaluate the car against competitors in the same segment.

Moreover, research has shown that owners tend to perceive their own car being of higher quality than non-owners eventually leading to statements more positive than the wider population. The data collected from the qualitative interview can therefore be seen to have quite low external validity. Nevertheless, the obtained quantitative data collected during the eye tracking has according to the authors' higher external validity. The main reason for this lies in the nature of eye tracking, meaning that the eyes will automatically track and glance towards areas and components of interest, reducing potential bias.

The main advantage using the pool of volunteers from the ergonomics department was the ability to perform the study within the limited time and budget. The pool of employees also allowed the authors to adjust the sample to include people from the different age segments of potential Volvo purchasers. The method of sending out different amounts of request within each age segment worked quite well (see table 15), it should however be mentioned that this method implies a risk since the response from the requested people is unknown, which eventually could lead to a poorly distributed sample.

Table 15. Distribution of Asked and respondent participants

Asked Participants Distribution			Respondents Distribution		
Born	**Amount**	**Percent**	**Born**	**Amount**	**Percent**
1980-1987	12	8	1980-1987	2	8
1970-1979	46	30	1970-1979	7	28
1960-1969	69	45	1960-1969	14	56
1949-1959	27	17	1949-1959	2	8

The mean age of a Volvo main driver is according to market research 52 years, whereas the mean age of the participants in this study is 47 years.

Although the employees have volunteered to participate in the pool, only 16 percent of the polled people participated in the actual study. This underlines the authors concerns stated in the planning of study, that using people outside Volvo would had require other means in terms of budget and time.

During the study the authors discussed distribution of gender. According to the market research is the main driver of Volvo distributed accordingly; 78 percent male and 22 percent women. However, since the market research do not include any details on who's actually making the decision in case of a purchase, the authors decided to include an equal distribution of men and women in the study.

Other topics that affect the representativeness of the sample are the income and education. These topics were decided not to be asked for, due to the sensitive nature of the questions. These data can therefore not be compared with the market research information.

The sample was handpicked to reduce potential bias, employees currently working with quality, interior and aesthetics was consequently removed from the poll. That said, there is still a probability that some of the participants earlier worked within the previously mentioned areas. There are some examples from the interviews when the participant states that he/she has been involved with the development of a specific area or component. Consequently were the authors needed to handle this issue by marking the interview as less valid.

The study performed includes evaluation of one specific model in Volvo Cars product offer. Although parts of the design can be recognized over several other models it should be noticed that the results from the study should be generalized to other models with prudence. The Center-stack is one of the areas on the newly launched XC90 where Volvo Cars drastically has changed the design and functionality. The buttons have been replaced with a touchscreen, theoretically lowering the visual complexity of the area. It is therefore assumed, in line with the previous discussion that complexity draws attention, that the time needed to evaluate this area is reduced. Consequently leading to changes in the overall glance pattern.

6.3.3 Internal validity- Design-Execution-Analysis

During the design, implementation and analysis of the study the authors have been working systematically to reduce bias and increase the internal validity. The aim has been to keep the complexity of the participants' tasks to a minimum order to reduce the possibility to variations. The semi-structured interview was designed with one structured part in the beginning, giving the same possibility for all participants to deliver statements without receiving any probing. Moreover was the manuscript designed to easily identify at which level of probing the various statements was delivered.

The variables such as blinks, cross-trough time and design of the Areas of interests has been the same during the analysis of the quantitative data, thus enables an equal probability for the participants and the AOI´s to be detected.

The analysis of the qualitative data has been reviewed independently by both of the authors and later discussed together to reduce individual bias.

6.3.4 The eye tracking systems impact on the reliability of the study

The reliability of the collected and interpreted data has been of great focus during all parts of this project. It has been underlined several times that eye tracking is a relatively immature technique, especially when it comes to interaction in three dimensions. Even though the study has been designed with respect to the systems limitations, only including interior analysis, some variables when using the system do have impact on the reliability.

The inconsistent performance from the calibration resulted in recognition rates ranging from 36-98 percent, with a mean value of 81. Since people have different anatomy and the system is manually calibrated, it is impossible to determine if a repeated study would lie in the same range, meaning the reliability of the study is reduced. Another issue is the loss of data, since the recognition rate mean value is 81 percent there are data consistently being un-captured.

The precision of the system also contains variables that are difficult to control. Since the participants were allowed to interact freely in the interior there is a risk of affect the position of the system resulting in an offset in the obtained data. It has also been obtained that during the evaluation of the vehicle there is a big difference in how fast the participants move their head. This will have consequences for the system's ability to register the environment and markers. It is possible that some participants felt insecure about the situation, resulting in rapid movements of the head.

Although the complexity of the study was reduced to match the performance of the eye tracking system, there are several indications enlightening that the complexity of the performed study is in the outer edge on what can be seen as acceptable in terms of reliability.

6.4 Answer to RQ3

The main study was conducted in order to find answer to research questions 3.

RQ 3: *What does a person look at most and first when evaluating the quality of a car?*

SRQ 3.1: *What is the reason to that people look where they look?*

SRQ 3.2: *Are people aware of where they look?*

The main study has showed that the areas that the participant focused most during the eye tracking evaluation were the Center Stack (14,2%) and the Steering Wheel (7,5%). These two areas has considerably higher attentions rates than the other areas included.

The area that the participants evaluate first is according to the eye tracking glance metrics TFG and the manual observation of the data, the Steering Wheel.

The answer to the subsequent research question 3.1 is answered by combining the eye-tracking data with the qualitative data collected during the interviews. The most evident finding is the correlation between the high attention ratio of the Center Stack and the comments during the interviews where the participants stated that the area was complex with a high number of components and functionalities. It is also concluded that components near the driver increases the chances for the area to be evaluated visually, due to the possibility for physical interaction.

The answer to subsequent research question 3.2 is that people are rather unaware of where they look. During the analyze of the eye tracking data and the statements given during the interviews it became evident that only 7 out of 25 participants have a direct correlation between where they looked most and what they stated.

7. CONCLUSIONS

The conclusion on the question if Dikablis essential eye tracking technology is mature enough to be used as an evaluation tool for the Perceived Quality of cars is Yes. The technology shows on many advantages in including potential customers into the evaluation of quality, but it is important to clearly state that there are many limitations forcing the system to be used in a certain manner.

Conclusions are made out from the limitations in the eye tracking system Dikablis Essential developed by Ergoneers Group and there may be improvements in other systems, both newer systems developed by Ergoneers Group but also other developers, that reduces problem areas and increase the potential for eye tracking used in the context of Perceived Quality evaluations.

The main limitation of Dikablis Essential is the low precision of the system, making it difficult to detailed Areas of Interest in the evaluation. Requested functionality expressed by Volvo Cars, such as following split-lines and focusing on smaller components on the car can't be done using this system.

With a more stable and precise system, there is also a possibility to cover more areas of the car with a higher validity in the result. In the study performed a high percentage of the glances are outside any of the AOI's, this percentage can be decreased with better stability given the possibility for more AOI's.

There are also limitations affecting the validity and reliability of a study performed with Dikablis Essential. The validity is mainly affected by the markers risk of drawing attention to certain areas and that the poor movability affects the behavior of the person using the system.

The reliability is mainly affected by the adjustment needed for each person, that the calibration is hard to perform in the exact same way person for person, and that the qualitative complementing collection of data always involves some kind of subjective judgment during the analysis.

The need of a complementing qualitative data collection method is established definite as the eye tracking data never can be done in a way to answer why a person is looking at a certain area. The conclusion is made that a well-planned personal interview is the best way to interact with the persons to understand the underlying reasons to the eye tracking data.

Even though a well-planned interview would provide additional information it is concluded that with the technology available today, there is no way to be entirely sure on why some areas get high attention in the eye tracking data without a subjective conclusion from the researchers.

7.1 Why an area is looked at more than other

The conclusions made by the authors in this thesis are that complexity is a main factor drawing attention and resulting in high attention ratio. High complexity of an area means that much time is needed to evaluate everything in order to interpret the area and to create an impression if the quality is good or bad. It is important to point out that the complexity and amount of time put on an area isn't correlated with if an area is good or bad.

A low complexity area can be looked at very fast with two possible reasons; one is that it is quickly interpreted as nothing stands out and it looks fine, or that it is quickly interpreted to look bad because it is something in the area which clearly stands out. Ones again qualitative information is shown to be needed in order to answer on why.

Another conclusion made from the discussion about quality, and especially Perceived Quality, is that it is hard to define and talk about for people. During the discussions people use the term "holistic" when commenting on quality and points out that everything needs to "harmonize" and "match" together.

These expressions goes well into the conclusion that an object is first perceived as bad when it clearly stands out from the expected harmonized holistic view. The difference in quality needs to be equal through all components and it isn't good if something stands out in either direction, as it will affect the perception of the other components. Comments have been made on material differentiating between components which are thought to be equal in quality (e.g. different gloss in chromium details).

An understanding of this conclusion is highlighted by the comments on the additional features, which is assumed to be less prioritized as they often does not harmonize as smoothly as with the rest of the interior, either by material selections or in shape.

The last conclusion made from the thesis is that people can be categorized according to what they tend to evaluate and comment on. The main categorizes is *Functionality*, *Material* and *Manufacturing and Assembling*, but the results shows on the possibility to further divide the potential customers into sub-categories.

8. RECOMMENDATIONS

It is recommended to look to the possibility of using eye tracking in the development process at Volvo Cars to include customers' view of quality together with qualitative interviews to better interpret the eye tracking data received.

In order to obtain data with higher validity and reliability it is recommended to look into other systems and developers, and to make a more in-depth research of the different systems on the market before making any bigger investments.

There are systems available on the market which is wireless with easier performed calibrations offering higher precision and stability in recording of eye tracking data. The technology is quite new (from 2009) and there are consistently new improvements in the developed eye tracking systems.

Whit a system offering high validity and reliability it is recommended to include real potential customers outside Volvo Cars to reduce the risk of bias and to increase the external validity of the result. Including outside customers is a more expensive sampling method and the results need to have good enough precision to justify such an investment.

A new system would mean that exterior evaluation can be performed and this is recommended as it is seen to better capture the attributes of the department of Geometry and Appearance as the focus on split-lines and geometry is easier obtained in exterior evaluation than in interior, as the complexity in the interior takes focus.

Together with an improved eye tracking system, it is also recommended to improve the interview manuscript to better capture the reason on why a person looks where it looks. Using mixed methods with quantitative and qualitative data collection methods is a new field for the two authors and there is more expertise to benefit from in improving the interviews.

8.1 Why an area is looked at more than other

Recommendations made when looking to the conclusions on why a person looks where it looks is that the holistic view needs to be worked on in order to maintain perception of premium quality.

The work of improving the quality of components in the interior will show less result if there is something else in the interior drawing attention. The quality is stated to be judged in a comparison context and if something is very well crafted and with a high quality material, everything else will be directly compared with that component. The same goes with a component badly crafted with low quality material.

A recommendation is to assure that all components have the similar level of quality and that controls are made through the departments to assure that the material and shapes of different components are harmonizing.

When the demands on the product total quality are raised the quality of each individual component needs to follow, and if a car is to be perceived as premium then each and every component of the car needs to be developed to match that premium look.

REFERENCES

alg. (2012). *Perceived Quality Study - Mainstream and Luxury Models*. Retrieved from

Aliaga, M., & Gunderson, B. (1999). *Interactive statistics*: Prentice Hall.

Baxter, M. Product Design: A Practical Guide to Systematic Methods of New Product Development, 1995: London: Chapman & Hall.

Bentivoglio, A. R., Bressman, S. B., Cassetta, E., Carretta, D., Tonali, P., & Albanese, A. (1997). Analysis of blink rate patterns in normal subjects. *Movement Disorders, 12*(6), 1028-1034.

Boyd, N. (2015). What is Internal Validity in Research? - Definition & Examples. Retrieved from http://study.com/academy/lesson/what-is-internal-validity-in-research-definition-examples.html

BusinessDictionary. (2015). Heatmap. Retrieved from http://www.businessdictionary.com/definition/heatmap.html

Byrne, M. D., Anderson, J. R., Douglass, S., & Matessa, M. (1999). *Eye tracking the visual search of click-down menus.* Paper presented at the Proceedings of the SIGCHI conference on Human Factors in Computing Systems.

Campbell, D., & Fiske, D. (1998). Convergent and discriminant validation by the multitrait-multimethods matrix'. *Personality, 56*, 162.

Caracelli, V. J., & Greene, J. C. (1993). Data analysis strategies for mixed-method evaluation designs. *Educational evaluation and policy analysis, 15*(2), 195-207.

Clodfelter, R., & Fowler, D. (2001). Do consumers' perceptions of product quality differ from objective measures of product quality. *Conway: University of Central Arkansas, Small Business Advancement National Center*.

Collewijn, H., & Kowler, E. (2008). The significance of microsaccades for vision and oculomotor control. *Journal of Vision, 8*(14), 20.

Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches*: Sage publications.

CSU. (2015). Validity. Retrieved from http://writing.colostate.edu/guides/page.cfm?pageid=1388&guideid=66

Custom-Analytics. (2014). What is a Heat Map? Retrieved from http://custom-analytics.thomsonreuterslifesciences.com/SpotfireWeb/Help/dxpwebclient/heat_what_is_a_heat_map.htm

Davies, D., & Dodd, J. (2002). Qualitative research and the question of rigor. *Qualitative health research, 12*(2), 279-289.

Denzin, N. K. (1978). The research act: A theoretical introduction to research methods: New York: McGraw-Hill.

Drever, E. (1995). *Using Semi-Structured Interviews in Small-Scale Research. A Teacher's Guide*: ERIC.

Duchowski, A. (2007). *Eye tracking methodology: Theory and practice* (Vol. 373): Springer Science & Business Media.

Dunn, K. (2000). Interviewing.

Ergoneers. (2011). D-Lab & D-Lab Control Plan. Measure. Analyse - User Manual.

Ergoneers. (2015). Our Mission Statement. Retrieved from http://www.ergoneers.com/en/ergoneers-group/mission/

Ergoneers.Group. (2011). *Dikablis - The Eye Tracking System, User Manual*.

Ergoneers.Group. (2012). *D-lab and D-lab Control - User Manual valid for D-lab 2.5*.

Ergoneers.Group. (2013). Heat Map. Retrieved from https://www.youtube.com/watch?v=n-ePvL9h6C8

Espejel, J., & Fandos, C. (2008). Perceived quality as a antecedent for buying intention of the olive oil from Bajo Aragón with protected designation of origin. *Esic Market, September-December*(131), 231-251.

Flowerdew, R., & Martin, D. (2005). *Methods in human geography: a guide for students doing a research project*: Pearson Education.

Foddy, W., & Foddy, W. H. (1994). *Constructing questions for interviews and questionnaires: theory and practice in social research*: Cambridge university press.

Garvin, D. A. (1984). What does product quality really mean. *Sloan management review, 26*(1).

GetEyeSmart. (2015). Eye Anatomy. Retrieved from http://www.geteyesmart.org/eyesmart/diseases/images/Eye-anatomy.jpg

Given, L. M. (2008). *The Sage encyclopedia of qualitative research methods*: Sage Publications.

Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The qualitative report, 8*(4), 597-607.

Goldberg, J. H., & Wichansky, A. M. (2002). Eye tracking in usability evaluation: A practitioner's guide. *To appear in: Hyönä*.

Greene, J. C. (2007). *Mixed methods in social inquiry* (Vol. 9): John Wiley & Sons.

Harrell, M. C., & Bradley, M. A. (2009). *Data collection methods. Semi-structured interviews and focus groups*. Retrieved from

Holzman, P. S., Proctor, L. R., Levy, D. L., Yasillo, N. J., Meltzer, H. Y., & Hurt, S. W. (1974). Eye-tracking dysfunctions in schizophrenic patients and their relatives. *Archives of general psychiatry, 31*(2), 143-151.

Horton, J., Macve, R., & Struyven, G. (2004). Qualitative research: experiences in using semi-structured interviews. *The real life guide to accounting research*, 339-357.

Jacob, R. J., & Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. *Mind, 2*(3), 4.

Jick, T. D. (1979). Mixing qualitative and quantitative methods: Triangulation in action. *Administrative science quarterly*, 602-611.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive psychology, 8*(4), 441-480.

Katz, J. (1983). A theory of qualitative methodology: The social system of analytic fieldwork. *Contemporary field research, 4*.

Kowler, E. (2011). Eye movements: The past 25years. *Vision research, 51*(13), 1457-1483.

Krauzlis, R. J., & Dill, N. (2002). Neural correlates of target choice for pursuit and saccades in the primate superior colliculus. *Neuron, 35*(2), 355-363.

Laerd-Dissertation. (2012). External validity. Retrieved from http://dissertation.laerd.com/external-validity.php

Lincoln, Y. S., & Guba, E. G. (1985). Naturalist inquiry. *Beverly Hills, CA: Sage*.

Lobe, B. (2008). *Integration of online research methods*: Faculty of Social Sciences.

Marshall, M. N. (1996). Sampling for qualitative research. *Family practice, 13*(6), 522-526.

Maxfield, J., Dew, P. M., Zhao, J., Juster, N., & Fitchie, M. (2002). *A virtual environment for aesthetic quality assessment of flexible assemblies in the automotive design process*. Retrieved from

Maynes, E. S. (1976). The concept and measurement of product quality *Household production and consumption* (pp. 529-584): NBER.

McQuarrie, E. F. (2011). *The market research toolbox: a concise guide for beginners*: Sage.

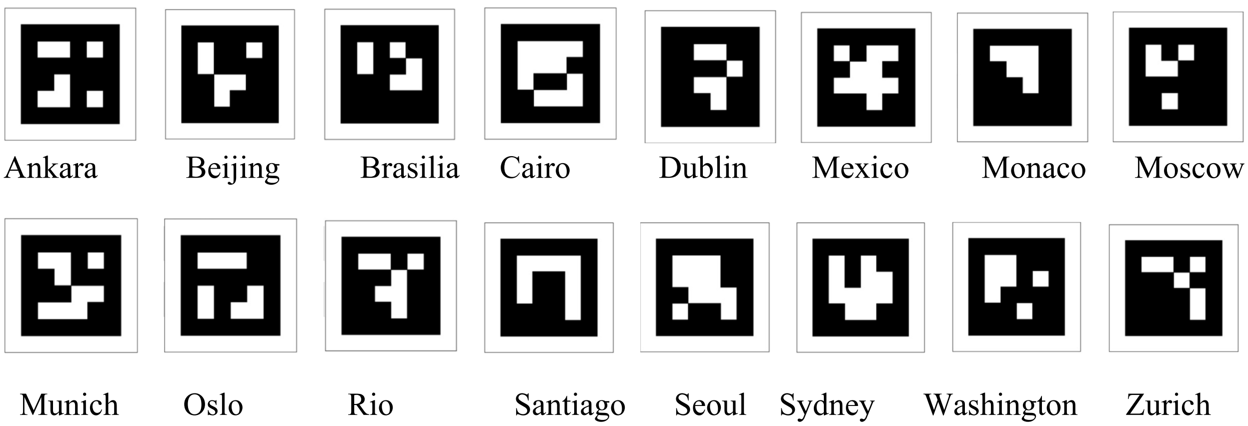
Messick, S. (1990). Validity of test interpretation and use. *ETS Research Report Series, 1990*(1), 1487-1495.

Milekic, S. (2004). The More You Look the More You Get: Intention-based Interface using Gaze-tracking. http://www.archimuse.com/mw2003/papers/milekic/milekic.html
Mitra, D., & Golder, P. N. (2006). How does objective quality affect perceived quality? Short-term effects, long-term effects, and asymmetries. *Marketing Science, 25*(3), 230-247.
Patton, M. Q. (1990). *Qualitative evaluation and research methods*: SAGE Publications, inc.
Paultan.org. (2013). Volvo V40. Retrieved from http://paultan.org/2013/11/18/video-polestar-teases-new-hot-volvo-v40/
Pollatsek, A., Rayner, K., & Collins, W. E. (1984). Integrating pictorial information across eye movements. *Journal of Experimental Psychology: General, 113*(3), 426.
Poole, A., & Ball, L. J. (2006). Eye tracking in HCI and usability research. *Encyclopedia of human computer interaction, 1*, 211-219.
Poole, A., Ball, L. J., & Phillips, P. (2005). In search of salience: A response-time and eye-movement analysis of bookmark recognition *People and Computers XVIII—Design for Life* (pp. 363-378): Springer.
PQ.Department. (2012). Strategidokument 2012.
Reid, T. N., MacDonald, E. F., & Du, P. (2013). Impact of Product Design Representation on Customer Judgment. *Journal of Mechanical Design, 135*(9), 091008.
Richardson, D. C., & Spivey, M. J. (2004). Eye tracking: Research areas and applications. *Encyclopedia of biomaterials and biomedical engineering*, 573-582.
Ritchie, J., Lewis, J., Nicholls, C. M., & Ormston, R. (2013). *Qualitative research practice: A guide for social science students and researchers*: Sage.
Rogelberg, S. G. (2002). *Handbook of research methods in industrial and organizational psychology*: John Wiley & Sons.
Rubin, H. J., & Rubin, I. S. (2011). *Qualitative interviewing: The art of hearing data*: Sage.
Sandelowski, M. (1995). Sample size in qualitative research. *Research in nursing & health, 18*(2), 179-183.
Sandelowski, M. (2000). Focus on research methods combining qualitative and quantitative sampling, data collection, and analysis techniques. *Research in nursing & health, 23*, 246-255.
Seale, C. (1999). Quality in qualitative research. *Qualitative inquiry, 5*(4), 465-478.
Segre, L. (2014). Human Eye Anatomy: Parts of the Eye. Retrieved from http://www.allaboutvision.com/resources/anatomy.htm
Shepard, L. A. (1997). The centrality of test use and consequences for test validity. *Educational Measurement: Issues and Practice, 16*(2), 5-24.
Shuttleworth, M. (2008). Validity and Reliability. https://explorable.com/validity-and-reliability
Silverman, D. (2010). *Qualitative research*: Sage.
Smith, S. (2013). Determining Sample Size: How to Ensure You Get the Correct Sample Size. http://www.qualtrics.com/blog/determining-sample-size/
Stenbacka, C. (2001). Qualitative research requires quality concepts of its own. *Management decision, 39*(7), 551-556.
Stylidis, K., Wickman, C., & Söderberg, R. (2015). Defining perceived quality in the automotive industry: an engineering approach.
Tobii. (2015). This is eye tracking. Retrieved from http://www.tobii.com/en/about-tobii/what-is-eye-tracking/
UCL. (2006). Blink and you miss it! http://www.ucl.ac.uk/media/library/blinking
van Raan, A. F. J. (2013). *Handbook of quantitative studies of science and technology*: Elsevier.
Volvo.Cars. (2014). Corporate Strategy. Retrieved from https://intranet.volvocars.net/volvo-car-group/corporate-strategy
Wainer, H., & Braun, H. I. (2013). *Test validity*: Routledge.

Weiss, R. S. (1995). *Learning from strangers: The art and method of qualitative interview studies*: Simon and Schuster.
Wikipedia. (2015). Eye Tracking. Retrieved from https://en.wikipedia.org/wiki/Eye_tracking
Wilson, C. (2013). *Interview techniques for UX practitioners: A user-centered design method*: Newnes.
Zeithaml, V. A. (1988). Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence. *The Journal of marketing*, 2-22.

APPENDIX

1. The 16 markers used in Dikablis Essential



2. Result from the marker recognition evaluation

Table over picture frames	Monaco	Sydney	Santiago	Dublin	Seoul	Cairo	Brasilia	Beijing
Mean Value (no./125)	87,813	76	74,9375	70,75	70,625	68,375	68,25	65,6875
Mean Value (%)	70,25	60,8	59,95	56,6	56,5	54,7	54,6	52,55
Rating	1	2	3	4	5	6	7	8
	Oslo	Moscow	Rio	Zurich	Mexico	Munich	Ankara	Washington
Mean Value (no./125)	65,438	65	62,8125	62,313	61,625	60,938	48,375	47,8125
Mean Value (%)	52,35	52	50,25	49,85	49,3	48,75	38,7	38,25
Rating	9	10	11	12	13	14	15	16

3. Result from the calibration evaluation

Calibration		SIZE 1			SIZE 2			SIZE 3		
	RR (%)	GTOT	GNUM	ARATIO	GTOT	GNUM	ARATIO	GTOT	GNUM	ARATIO
Wide										
Test 1	97,00	4,92	32,00	13,82	5,80	31,00	16,29	5,80	31,00	16,29
Test 2	95,00	0,12	3,00	0,31	1,92	18,00	4,92	23,44	52,00	60,10
Test 3	96,00	0,00	0,00	0,00	0,00	0,00	0,00	0,04	1,00	0,11
Test 4	93,00	0,00	0,00	0,00	0,84	9,00	2,30	20,40	58,00	55,86
Test 5	95,00	0,00	0,00	0,00	0,04	1,00	0,12	0,28	4,00	0,82
	Mean	1,01	7,00	2,83	1,72	11,80	4,73	9,99	29,20	26,64
Narrow										
Test 1	96,00	7,20	19,00	21,82	15,28	27,00	46,30	25,08	20,00	76,00
Test 2	95,00	5,24	21,00	17,54	14,80	37,00	49,53	16,64	28,00	55,69
Test 3	97,00	2,52	11,00	7,29	9,92	22,00	28,70	24,88	26,00	71,99
Test 4	95,00	5,68	25,00	18,86	17,64	25,00	58,57	23,56	10,00	78,22
Test 5	96,00	17,76	35,00	62,10	21,32	12,00	74,55	22,00	15,00	76,92
	Mean	6,21	14,60	20,25	9,73	18,20	31,25	16,21	24,50	**71,76**

4. Glance metrics from exterior and interior evaluation in Pilot Study

	GTOT	GNUM	GMEAN	ARATIO	GTOT	GNUM	GMEAN	ARATIO	GTOT	GNUM	GMEAN	ARATIO	GTOT	GNUM	GMEAN	ARATIO
Exterior	Left Front Light				Right Rear View Mirror				Right Back Light				Grill			
TP 1	¤	¤	¤	¤	2,52	10	0,25	1,64	2,72	4	0,68	1,77	NAN	NAN	NAN	NAN
TP 2	1,12	10	0,11	0,41	4,72	17	0,28	1,72	3,08	14	0,22	1,12	2,88	7	0,41	1,05
TP 3	2	13	0,15	0,97	2,04	5	0,41	0,98	0,2	1	0,20	0,10	9,60	18	0,53	4,63
TP 4	2,68	4	0,67	2,19	3,32	10	0,33	2,71	0,04	1	0,04	0,03	4,08	15	0,27	3,33
TP 5	NAN	NAN	NAN	NAN	3,88	8	0,49	2,56	NAN	NAN	NAN	NAN	0,04	1	0,04	0,03
MEAN	1,45	6,75	0,23	0,88975	3,296	10	0,35	1,92	1,208	4	0,23	0,604	3,32	8,2	0,25	1,81
	GTOT	GNUM	GMEAN	ARATIO	GTOT	GNUM	GMEAN	ARATIO	GTOT	GNUM	GMEAN	ARATIO	GTOT	GNUM	GMEAN	ARATIO
Exterior	Left Rear View Mirror				Plenum				Right C-pillar				Left C-Pillar			
TP 1	0,52	1	0,52	0,34	1,28	5	0,26	0,83	NAN	NAN	NAN	NAN	1,12	5	0,22	0,73
TP 2	4,64	16	0,29	1,69	1,44	10	0,14	0,52	0,52	3	0,17	0,19	0,56	3	0,19	0,20
TP 3	NAN	NAN	NAN	NAN	¤	¤	¤	¤	0,40	1	0,40	0,19	0,04	1	0,04	0,02
TP 4	NAN	NAN	NAN	NAN	0,32	1	0,32	0,26	0,20	1	0,20	0,16	¤	¤	¤	¤
TP 5	NAN	NAN	NAN	NAN	0,88	9	0,10	0,58	NAN	NAN	NAN	NAN	0,08	1	0,08	0,05
MEAN	1,03	3,4	0,16	0,41	0,98	6,25	0,20	0,55	0,22	1	0,15	0,11	0,45	2,5	0,13	0,25
	GTOT	GNUM	GMEAN	ARATIO	GTOT	GNUM	GMEAN	ARATIO	GTOT	GNUM	GMEAN	ARATIO	GTOT	GNUM	GMEAN	ARATIO
Exterior	Right B-Pillar				Lower Back Center				Upper Back Center				DLO passenger			
TP 1	0,04	1	0,04	0,03	5,04	12	0,42	3,28	5,48	14	0,39	3,57	¤	¤	¤	¤
TP 2	3,74	9	0,42	1,25	7,60	12	0,63	2,76	1,68	6	0,28	0,61	¤	¤	¤	¤
TP 3	1,16	1	1,16	0,56	4,20	15	0,28	2,03	1,36	5	0,27	0,66	¤	¤	¤	¤
TP 4	5,88	11	0,53	4,80	4,96	8	0,62	4,05	1,00	3	0,33	0,82	¤	¤	¤	¤
TP 5	0,64	2	0,32	0,42	5,64	13	0,43	3,73	0,52	3	0,17	0,34	0,16	2	0,08	0,106
MEAN	2,29	4,8	0,49	1,41	5,49	12	0,48	3,17	2,01	6,2	0,29	1,20	0,16	2	0,08	0,106
	GTOT	GNUM	GMEAN	ARATIO	GTOT	GNUM	GMEAN	ARATIO	GTOT	GNUM	GMEAN	ARATIO	GTOT	GNUM	GMEAN	ARATIO
Exterior	Left back light				Left B-pillar				DLO Driver				Right Front Light			
TP 1	0,08	2	0,04	0,05	0,60	2	0,30	0,39	¤	¤	¤	¤	0,40	4	0,10	0,26
TP 2	0,24	1	0,24	0,09	6,24	22	0,28	2,27	¤	¤	¤	¤	0,72	7	0,10	0,26
TP 3	0,64	4	0,16	0,31	1,88	10	0,19	0,91	¤	¤	¤	¤	0,04	1	0,04	0,02
TP 4	0,88	2	0,44	0,72	1,32	3	0,44	1,08	0,92	7	0,13	0,75	¤	¤	¤	¤
TP 5	0,68	1	0,68	0,45	5,52	7	0,79	3,65	¤	¤	¤	¤	1,24	4	0,31	0,82
MEAN	0,50	2	0,31	0,32	3,11	8,8	0,40	1,66	0,92	7	0,13	0,75	0,60	4	0,14	0,34

	GTOT	GNUM	GMEAN	ARATIO	GTOT	GNUM	GMEAN	ARATIO	GTOT	GNUM	GMEAN	ARATIO	GTOT	GNUM	GMEAN	ARATIO
Interior	Steering Wheel				Ceiling				Left IP				Right A-Pillar			
TP 1	7,72	29	0,27	6,24	NAN	NAN	NAN	NAN	8,16	9	0,91	6,60	0,24	2	0,12	0,19
TP 2	11,16	22	0,51	9,24	2,24	3	0,7	1,85	1,24	6	0,21	1,03	0,92	5	0,18	0,76
TP 3	12,72	47	0,27	7,00	5,64	13	0,4	3,10	18,16	73	0,25	9,99	¤	¤	¤	¤
TP 4	5,48	15	0,37	5,55	5,76	5	1,2	5,84	7,52	14	0,54	7,62	NAN	NAN	NAN	NAN
TP 5	6,76	32	0,21	3,06	6,04	10	0,6	2,74	12,72	46	0,28	5,76	0,56	3	0,19	0,25
MEAN	8,768	29	0,32	6,22	3,936	6,2	0,6	2,71	9,56	29,6	0,44	6,20	0,43	2,50	0,12	0,30
	GTOT	GNUM	GMEAN	ARATIO	GTOT	GNUM	GMEAN	ARATIO	GTOT	GNUM	GMEAN	ARATIO	GTOT	GNUM	GMEAN	ARATIO
Interior	Right Air Vent				Right IP				Left A-Pillar				Glove Compartment			
TP 1	0,24	2	0,12	0,19	0,04	1	0,04	0,03	NAN	NAN	NAN	NAN	0,04	1	0,04	0,03
TP 2	1,08	4	0,27	0,89	NAN	NAN	NAN	NAN	NAN	NAN	NAN	NAN	0,68	6	0,11	0,56
TP 3	¤	¤	¤	¤	¤	¤	¤	¤	0,32	3	0,11	0,18	¤	¤	¤	¤
TP 4	0,32	2	0,16	0,32	0,40	4	0,1	0,41	0,36	4	0,09	0,37	0,08	2	0,04	0,08
TP 5	NAN	NAN	NAN	NAN	0,88	3	0,29	0,40	NAN	NAN	NAN	NAN	0,16	1	0,16	0,07
MEAN	0,41	2	0,14	0,35	0,33	2	0,11	0,21	0,14	1,4	0,04	0,11	0,24	2,5	0,09	0,19
	GTOT	GNUM	GMEAN	ARATIO	GTOT	GNUM	GMEAN	ARATIO	GTOT	GNUM	GMEAN	ARATIO				
Interior	Stereo				Center Console				Left Air Vent							
TP 1	11,20	13	0,86	9,06	¤	¤	¤	¤	0,56	8	0,07	0,45				
TP 2	1,28	3	0,43	1,06	¤	¤	¤	¤	0,44	2	0,22	0,36				
TP 3	10,24	54	0,19	5,63	1,36	1	1,36	0,75	2,96	2	1,48	1,63				
TP 4	2,80	20	0,14	2,84	¤	¤	¤	¤	2,12	13	0,16	2,15				
TP 5	5,12	26	0,20	2,32	3,04	4	0,76	1,38	4,52	16	0,28	2,05				
MEAN	6,13	23,2	0,36	4,18	2,20	2,5	1,06	1,06	2,12	8,2	0,44	1,33				

5. The Study Guide

Test Person: ....

Date: ....

Study Guide

Eye tracking study conducted at Volvo Cars in cooperation with the department of Perceived Quality.

Introduction

I would first like to thank you for participating in this study. Me and Anton is performing a master thesis in cooperation with Volvo Cars where we are investigating the possibility of using eye tracking technology as a visual evaluation of cars.

I would first like to know a bit more about your position and relation to Volvo Cars and your habits of driving.

Main Question	Subsequent question 1	Subsequent Question 2	
At which department are you currently working at Volvo Cars?			
What are your areas of responsibility in you work?			
For how long have you been working at Volo Cars?			
Do you currently own a car?	If yes: What type of car?		
	What kind of cars have you earlier owned?		
	What are you looking at when purchasing a new car?	(If a lot of things, ask them to please mention the most important)	
	What is most important if you are to choose between interior and exterior?		
How often are you driving?	(Give examples: ones a day, ones a week, ones a month)		

Test Person: ______________________

Date: ______________

- We will now move on with the evaluation of the car using eye tracking technology. Anton will assist you in getting the glasses adjusted and calibrated. Please notice that there are markers like these (showing examples) placed in the interior of the car and we would like you to try not to focus on them but just focusing on evaluating the interior quality.
- Assure that it feels okay for the test person with the glasses on when the calibration is done.
- I would now like you to evaluate the quality of the car in front of you. The evaluation is totally free and you are the one which decides what's important to evaluate when evaluating the interior quality. There is no right or wrong!
- I will afterwards ask you to grade the overall interior quality on a scale from 1-10.
- Feel free to move around in the car and there are no restrictions in how you can move your head, we would like to encourage you to follow along with the head and not only looking with your eyes.
- The evaluation starts directly from the moment you are gripping the door handle. Is everything clear?
- If yes, please do start when Anton is telling you to.

(When the free evaluation is stopped by the test persons, he/she is asked to stay seated and the interview is performed within the car. Answers in the initial questions is rated as probing level 3)

Main Question	**Subsequent question 1**	**Subsequent question 2**
We can start by you grading the total interior quality on a scale from 1-10.	On a scale from 1-10. How good do you assess the interior quality of the car to be? Answer: ________	
What are you looking at most when you are to evaluate the interior of a car?	If no area is mentioned: Is there any certain area in the car which you are looking at more than other areas?	Why?
What are you looking at first when evaluating the interior quality of a car?	Why?	
What areas do you think is of greatest importance when evaluating the interior quality of a car?	Why?	

Test Person: ______________________ Date: ______________

<table>
<tr><td>Is there any certain area/component/part which you think is a good example of good quality?</td><td>Why?</td><td></td><td></td></tr>
<tr><td>Is there any certain area/component/part which you think is a good example of bad quality?</td><td>Why?</td><td></td><td></td></tr>
<tr><td colspan="3">(After the initial questions, each area mentioned is more discussed. Answers here is rated as probing level 2)</td><td></td></tr>
<tr><td>If we are to focus on a certain area. What are your perceptions of the quality in this area? (area according to what is mentioned as important in the initial questions)</td><td>Why?</td><td></td><td></td></tr>
<tr><td>Could you give some examples of good quality in the area?</td><td>Why is it good?</td><td></td><td></td></tr>
<tr><td>Could you give some examples of bad quality in the area?</td><td>Why is it bad?</td><td></td><td></td></tr>
<tr><td colspan="3">(When all the areas mentioned as important is discussed)

Now I would only like to ask some finishing questions to sum up the interview.

• Do you think that the overall quality of the car is standing up to your expectations?

• Would you like to change your grading on the scale of 1-10 when you have looked through a bit more?

• What do you people in general perceives as important areas in a car when evaluating the interior quality?

(The interview is done!)

Thanks a lot for your participation. We will come back to all participants with the results obtained through the study.</td><td></td></tr>
</table>

x

Test Person: ______________________________ Date: ______________

(Picture to mark the attention of the test person while evaluating the interior)

(This is to be used in the interview to highlight importance and to focus the questions according to the attention)

6. The heat map study

7. The data over where people look most

TP	The AOI D-lab calculated the TP looked at the most	ARATIO	Attribute the TP expressed he/she looked at the most	Area the TP expressed he/she looked at the most
2	Steering Wheel	13,492	Surface material	Seat Door Center stack
3	Center Stack	21,871	Chair adjustments	Steering Wheel Gearshift
4	Center Stack	10,675	Layout/design Stitches	Steering Wheel Center Stack
5	Center Stack	10,62	Feel (things you reach)	Steering Wheel
6	Center Stack	14,536	Fit and finish	Not disclosed
7	Ceiling	4,273	Controls and buttons Solidity	Center Stack
8	Center Stack	6,547	Layout/design Stitches Controls and buttons Solidity	Seat Steering Wheel
9	Center Stack	9,202	Feel in material	Seat Steering Wheel
10	Center Stack	16,043	Fit and finish	Driver Environment
11	Center Stack	31,201	Fit and finish Solidity Matching material	Instrument Panel Dorr Panels
12	Center Stack	20,938	Surface Material	Steering Wheel Center Stack
13	Center Stack	19,142	Overlook	Center Stack Instrument panel Seat
14	Center Stack	16,645	Feel in buttons	Steering Wheel
15	Back	5,906	Feel in material	Seat Dimmer Steering wheel
16	Center Stack	15,203	Controls and buttons	Seat Door handle and panels
17	Center Stack	16,667	Simplicity	Driver Environment
18	Center Stack	21,477	Fit and finish Feel in buttons	Stereo
19	Center Stack	23,855	Fit and finish	Instrument Panel

20	Center Stack	19,114	Feel in material	Steering Wheel
21	Center Stack	13,7	Seating comfort	Seat
22	Center Stack	9,435	Fit and Finnish Simplicity	Seat
23	Center Stack	19,794	Fit and Finnish	Not disclosed
24	Center Stack	8,019	Surface material	Instrument Panel Center Stack
25	Center Stack	8,3	Overlook Fit and finish	Steering Wheel Center stack
26	Center Stack	7,417	Overlook Choice of color	Seat

8. Evaluation time and Validity Index

TP	Total Time (TT)	Recognition Rate (RR)	Marker Validity (MV)	Eye Validity (EV)
1	03:08	60	72,01	75,90
2	05:33	57	77,86	80,77
3	02:38	36	91,87	71,87
4	01:54	97	75,88	60,94
5	04:04	75	87,59	82,58
6	02:33	41	60,28	72,61
7	02:00	77	43,68	46,94
8	04:01	63	48,19	83,05
9	03:04	88	72,97	78,11
10	10:37	96	89,36	94,44
11	02:24	93	71,22	75,80
12	10.44	94	83,63	94,80
13	02:01	79	80,38	72,58
14	04:10	82	55,51	87,01
15	03:43	97	81,24	85,51
16	03.12	93	77,51	67,10
17	22:06	90	85,25	95,44
18	04:17	94	77,63	77,21
19	02:38	91	75,97	80,18
20	02.45	94	67,50	94,74
21	01:43	93	57,83	69,94
22	09.08	96	88,62	89,37
23	03:00	71	69,64	71,24
24	01:58	98	74,28	51,60
25	03:14	74	69,96	84,17